Pearsons Correlation

$$w_{uv} = \frac{\sum_{i \in I_u \cap I_v} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u} (r_{ui} - \bar{r}_u)^2} \cdot \sqrt{\sum_{i \in I_v} (r_{vi} - \bar{r}_v)^2}}$$

similarity user u-v $[-1, 1]$

I_u: items rated by u

$$\cos(u, v) = \frac{\langle u, v \rangle}{\|u\| \cdot \|v\|} = w_{uv}$$

cosine similarity

User-User Collaborative Filtering

$$s(u,i) = \bar{r}_u + \frac{\sum_{v \in N(u)} w_{uv} (r_{vi} - \bar{r}_v)}{\sum_{v \in N(u)} |w_{uv}|}$$

$s(u,i)$: predicted rating user u - item i

$\bar{r}_u$: average rating u

$N(u)$: neighbouring users of u - highest w_{uv}

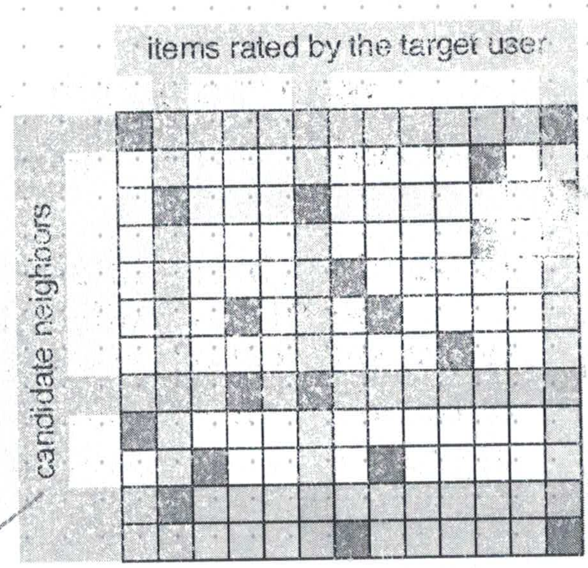
recommend item with highest $s(u,i)$

slide 35 optimization

optimizations for U-U CF

Optimization 1: neighborhood creation

- consider the **target user**
- due to *sparsity*, has **rated few items**
- to compute similarity with other users, it suffices to **look among users that have rated at least one of those items**
  - why? if no common rated item exists, the similarity is zero



compute similarity for these

Problems U-U CF

all pairwise user similarities $O(\underbrace{n}_{\#\text{items}}\underbrace{m^2}_{\#\text{users}})$ most complex

few ratings / sparse, users have no common items → no similarity u-u

can't precompute neighbourhood - new ratings change $N(u)$

Shilling attack: Inject similar users, manipulate recommendations

Item-Item Collaborative Filtering

$$w_{ij} = \frac{\sum_{u \in U_i \cap U_j} (r_{ui} - \bar{r}_i) \cdot (r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u \in U_i} (r_{ui} - \bar{r}_i)^2} \cdot \sqrt{\sum_{u \in U_j} (r_{uj} - \bar{r}_j)^2}}$$

similarity item i-j

users who have rated item i

score user-item

$$s(u,i) = \bar{r}_i + \frac{\sum_{j \in N(i)} w_{ij} (r_{uj} - \bar{r}_j)}{\sum_{j \in N(i)} |w_{ij}|}$$

average rating of item i

neighbour items of i - high w_{ij}
standard: take 20 neighbours

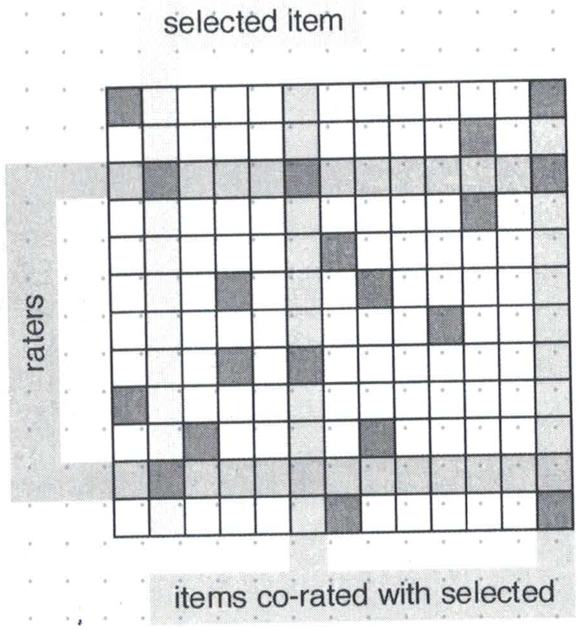
computing all pairwise item similarities $O(n^2 m)$ $n \ll m$!!

(5)

optimizations to I-I CF

- Optimization 1: similarities computation
  - exploit the sparsity of the ratings

- for each item, called **selected**
- identify its **raters**
- identify other **items co-rated** by those users
- compute **similarity** between selected and each of the co-rated items



[2003 IEEE G. Linden et al] *Amazon.com Recommendations*

(6)

Strength of I-I CF

more efficient than U-U CF

item-item similarities more stable - less items with more ratings

Weakness of I-I CF

recommendations obvious - items similar to ones already rated by u

Also consider novelty / serendipity / diversity

Feedback

⑦

Explicit: star rating, thumbs up
strong, but rare

Implicit: purchase, click, view
more abundant, but ambiguous
1: there was interaction, 0: no interaction interaction matrix
number of clicks / views

normalize user vectors (not centering) - compute cosine similarity

---

Cold Start

⑧

bad: new users / items / system

New User: non-personalized: popular, trending, age, gender
implicit feedback
learn users friends: social network, ~~reffenoo~~ referral

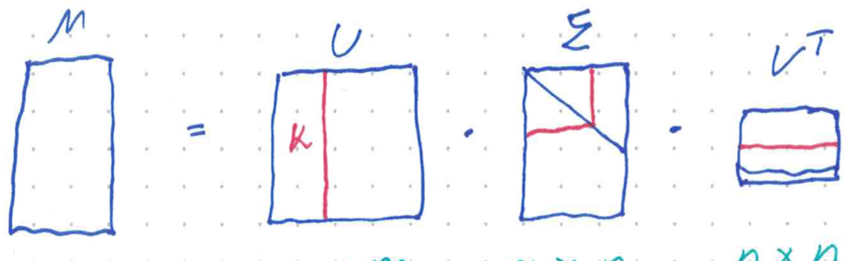
New Item: content based / descriptions
recommend to tolerant users / early adopters

New System: worst case
data from outside sources
knowledge based: ask for user preferences

Model based

compact representation
efficiency - faster
effectiveness - removes noise, less overfitting

Singular Value Decomposition - problem: sparse matrix / imputation



$M = U \cdot \Sigma \cdot V^T$

$m \times n$ $\quad$ $m \times m$ orthogonal $\quad$ $m \times n$ $\quad$ $n \times n$ orthogonal - $V^T = V^{-1}$

Truncate to k largest for $\hat{M}$

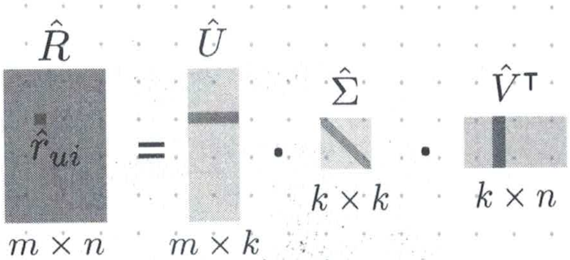
values < min(m,n) - singular

$$\underset{m\times n}{\hat{R}} = \underset{m\times k}{\hat{U}} \cdot \underset{k\times k}{\hat{\Sigma}} \cdot \underset{k\times n}{\hat{V}^T}$$

slide 13 SVD (10)

approximate ratings

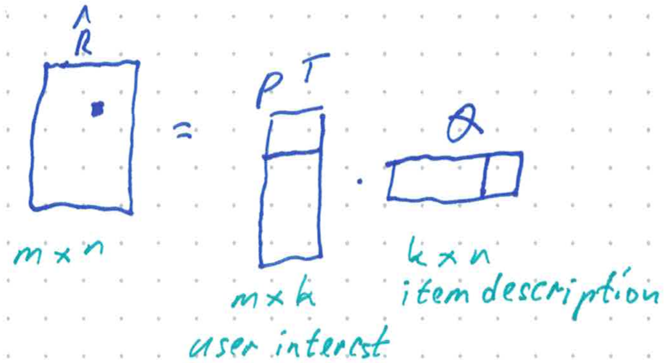
- the approximate/predicted rating of user u to item i is:



$$\hat{r}_{ui} = \sum_f \hat{U}_{uf} \cdot \hat{\Sigma}_{ff} \cdot \hat{V}_{if}$$

- sum over all features
- for each feature, multiply the user's interest $\hat{U}_{uf}$ with the item's description $\hat{V}_{if}$ and scale by feature significance $\hat{\Sigma}_{ff}$

Matrix Factorization



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

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

overall bias, bias user, bias item

Find P, Q to minimize J through stochastic gradient descent

$$J = \frac{1}{|R|} \sum_{r_{ui} \in R} (r_{ui} - p_u^T q_i)^2 + \lambda (\|P\|^2 + \|Q\|^2)$$

squared error — regularize P, Q

Stochastic Gradient Descent

$$\theta \leftarrow \theta - \eta \cdot \frac{\partial J^{(i)}(\theta)}{\partial \theta}$$

Algo:
- shuffle ratings R
- random init P, Q
- repeat
  - foreach $r_{ui} \in K$:
    - $e_{ui} \leftarrow r_{ui} - p_u^T q_i$
    - $p_u \leftarrow p_u + \eta (e_{ui} \cdot q_i - \lambda p_u)$
    - $q_i \leftarrow q_i + \eta (e_{ui} \cdot p_u - \lambda q_i)$
- until convergence

calculate bias as well

$b_u \leftarrow b_u + \eta \cdot (e_{ui} - \lambda b_u)$

$b_i \leftarrow b_i + \eta \cdot (e_{ui} - \lambda b_i)$

SVD++

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T \left(p_u + |N(u)|^{-\frac{1}{2}} \sum_{j \in N(u)} y_j \right)$$

bias — item model — user model — implicit feedback — items u clicked

---

Factorization Machines

x^{ui}: 0 1 0 1000 0,3 0,3 0,3 0 13 0001

user — movie — other movies — Time — last movie rated

$$\hat{r}_{ui} = \mu + \sum_p w_p x_p^{ui} + \sum_p \sum_{p' \neq p} w_{pp'} x_p^{ui} x_{p'}^{ui}$$

rating — bias — linear regression 1st order — pairwise relationships 2nd order

$w_{pp'} = v_p^T v_{p'}$ latent vector

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

$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T \left(p_u + |N(u)|^{-\frac{1}{2}} \sum_{j = N(u)} y_j \right)$ SVD++

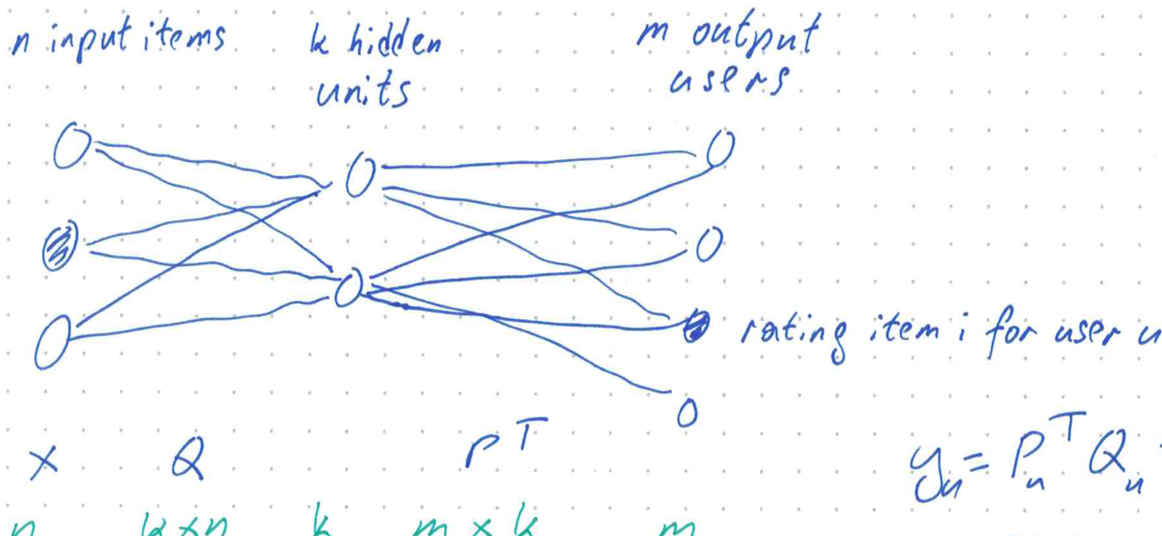
Sparse Linear Methods

$$\hat{R} = R\,Q^T Y$$

$$\underbrace{\hat{R}}_{m\times n} = \underbrace{K}_{m\times n} \cdot \underbrace{Q^T}_{n\times k} \cdot \underbrace{Y}_{k\times n}$$

$$\underbrace{\hat{r}_{ui} = q_i^T}_{\text{item model}} \; \underbrace{\sum_{j\in I_u} r_{uj}\cdot y_j}_{\text{user model}}$$

Neural Network vs. Matrix Factorization



$$y_u = P_u^T Q_u = \hat{r}_{ui}$$

$$y = P^T Q x$$

Weighted Matrix Factorization

$w_{ui} = 1 + \alpha c_{ui}$ $\quad$ c_{ui} = observed count

$w_{ui} = w \frac{f_i^a}{\sum_j f_j^a}$ $\quad$ missing feedback $\quad$ f_i = popularity of item i

Cost function

$$J = \frac{1}{n \cdot m} \sum_{u,i} w_{ui} e_{ui}^2 + \lambda (\|P\|^2 + \|Q\|^2)$$

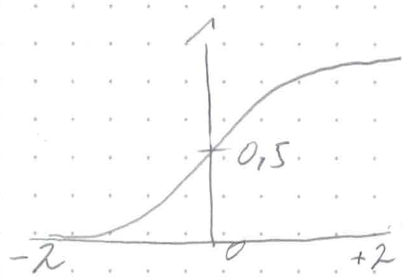
high confidence → bigger error

Bayesian Personalized Ranking

(u, i, j) $\quad$ $i > j$ $\quad$ i liked more than j

$\sigma(\hat{x}_{uij}) = P(i > j)$

$\hat{x}_{uij} = \hat{r}_{ui} - \hat{r}_{uj}$



pair-wise learning – better ranking accuracy

Content Based Recommenders

$$tf_{t,d} = \begin{cases} \log_{10}(1+\overbrace{f_{t,d}}^{\text{frequency of term in document}}) \\ 0 \end{cases}$$

term frequency

number of documents

$$idf_t = \log_{10}\left(\frac{N}{df_t}\right)$$

inverse document frequency

docs containing t

$$w_{t,d} = tf_{t,d} \cdot idf_t = \log_{10}(1+f_{t,d}) \cdot \log_{10}\left(\frac{N}{df_t}\right)$$

$$\text{document} = \langle w_{1d}, w_{2d}, w_{3d}, \dots w_{nd} \rangle$$

number of terms in document d

Similarity in Vector Space Model

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|q||d|} = \frac{q}{|q|} \cdot \frac{d}{|d|} = \frac{\sum_{i=1}^{|V|} q_i d_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2} \sqrt{\sum_{i=1}^{|V|} d_i^2}}$$

tf-idf of term i in past rated item

tf-idf of term i w_i in target item

$$\hat{r}_{uq} = \frac{\sum_{d \in N(q;u)} \cos(q,d) \cdot r_{ud}}{\sum_{d \in N(q;u)} \cos(q,d)}$$

User Profile - Relevance Feedback

$$u = \alpha \cdot u_0 + \beta \cdot \frac{1}{|D^+|} \sum_{d^+ \in D^+} d^+ - \gamma \cdot \frac{1}{|D^-|} \sum_{d^- \in D^-} d^-$$

vector baseline vector

compute cosine similarity of u to target item

Benefits: no other user profiles needed / no cold start problem
transparency - recommendation explained by past user likes

Drawbacks: need good features
no novelty / filter bubble.
some user feedback necessary

Rating Accuracy

Mean Absolute Error: $\frac{1}{|R|} \sum_{r_{ui} \in R} |e_{ui}|$ $e_{ui} = r_{ui} - \hat{r}_{ui}$

Root Mean Squared Error: $\sqrt{\frac{1}{|R|} \sum_{r_{ui} \in R} e_{ui}^2}$ penalizes big errors

Classification Accuracy

Precision: $\frac{TP}{TP + FP}$ only recommend relevant items Recall: $\frac{TP}{TP + FN}$ find all relevant items

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} = \frac{2}{\frac{1}{P} + \frac{1}{R}}$$

Ranking Accuracy

Precision @ k
how many TP until k compared to k max ↓

Recall @ k
how many TP of all found until k

Average Precision
count Precision at increases

Normalized Discounted Cumulative Gain

relevance of i at rank j to user u

$$DCG@k = \sum_{j=1}^{k} \frac{r_{u,i_j}}{\log(j+1)}$$

$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

Offline Experiment existing Dataset / Ground Truth, no actual user

User Studies Focus Groups, Usability Lab test - perform set of tasks

Online Experiment large scale on live system A/B Test, real users
risky, but strongest evidence: real user - real system, after offline test

Item Coverage long tail recommendations

User Coverage cold start users

Trust provide explanations for recommendation, hard to measu

Diversity intra-list diversity, distance between two recommendations

Novelty dissimilarity to past recommendations

Serendipity element of surprise

User Test: Between Subjects: different rec. per user, for long term, more data
Within Subjects: several rec. per user, allows comparative questions

Significance Testing: p value $< 0{,}01$

Conversational Recommender Systems (25)

Task oriented, multi turn, elicit preferences, provide explanation

Form based vs. Natural language - written or spoken
structured unstructured
system driven

System driven vs. User Driven vs Hybrid
system active - user passive rare most common/flexible

Components/Underlying Knowledge
- Item Catalogue
- User Intent
- User Model
- Dialogue States
- Background Knowledge: item related, Corpora, History

Classification of Goup Recommender Systems (26)

Group Type: established, occasional, random
Individual Preferences: known / unknown prior to group recommenda
Recommendation consumption: Rec. experienced / presented to group
Behaviour of Group: passive / active negotiation
Recommendation Type: single item / sequence of items

Aggregation Strategies

Additive: sum items over individuals

Multiplicative

Bordo Count: bottom zero - count up, tie: split points, sum up

Copeland Rule

Approval voting: count not strongly disliked / over threshold

Least Misery: minimum of individual ratings, then choose max valu

Most Pleasure: maximum of individual ratings

Average without Misery: only average of rec. where all above threshold

Social Choice Theories

Majority: Plurality

Consensus: Average, Average without misery

Borderline: Dictatorship, Least Misery, Most Pleasure

Objective AI ⟷ Subjective AI

Computer Vision   NLP   Recommendation

far from human   close to human

exact answers   no absolute Answer

Too many options → Beam Search

Item IDs can eliminate hallucinations

IDs should be distinguishable

Similar items similar id

dissimilar items dissimilar ID

Random ID: item<73><91>, item <73><29> differen items same token

Title ID: Inside Out, Inside Job different movies share same token

Independent ID: Item<1364>, Item<6321> too many out of vocabulary token

Consequences of Unfairness (29)

Information Asymetry

Matthew Effect

Echo Chambers

Fairness: Individual Fairness – Group Fairness

Harm: Distributional

Representational

Fairness for: Consumers

Providers

Subjects

(30)

Fairness Evaluation Dimensions

Statistical Parity equal prob. positively classified

Equal Opportunity True positive Rate is same

Equalized Odds: True positive and False positive is same

Overall Accuracy Equality same Accuracy ~~of~~ over Groups

Counterfactual Fairness output consistent when attribute changed