Where are we?

- Refinement/Personalization of results
- Study techniques of

  Recommender systems

  - Content filtering
    - Applying content filtering to search
  - Collaborative filtering
    - Nearest neighbor methods
      - Applying nearest neighbor method to search
    - Matrix factorization methods

  Detour to Latent Semantic indexing

1

Matrix factorization motivation

- Matrix representation
  - users X items
  - documents X terms
- Discover/use latent factors
  - attributes, topics, features
- Factor matrices to uncover latent factors
- Don't know what latent factors represent
  - can conjecture
- For recommenders, matrix has holes
  - use factorization to fill in

2

Matrix factorization for Collaborative Filtering

- Give ratings matrix R: M users X N items
  - R has holes- R_{ij} with no value
- Want to fill in holes => predict ratings
- Idea: decompose R:

$$R=PQ^T$$

  - P is M X f; Q is N X f
  - f dimensions are latent factors
    - no interpretation but can add one
  - must choose f

3

How does decomposition help?

- estimate P and Q, leaving no holes
- get estimate of R as $R_f = PQ^T$
  - R_f has holes of R filled in
- Several methods for estimation, e.g.
  - Gradient descent
  - Stochastic gradient descent
    - Koren et al. *Matrix Factorization Techniques for Recommender Systems*, IEEE Computer, Aug 2009
  - Least squares based calculations
    - Bell et al *Modeling Relat'ships at Multiple Scales to Improve Accuracy of Large Recom. Sys.*, KDD Aug 2007.

4

Optimization

- Minimize least squares error:

err(P,Q) is defined as

$$\sum_{(u,i) \text{ in } K}(R_{(u,i)} - (PQ^T)_{(u,i)})^2$$

for K the set of (u,i) for which $R_{(u,i)}$ has a value

5

Simple Step: Gradient Descent

- Minimize for one element change:
  - choose one element of P or one element of Q to vary, say $P_{(r,s)}$
    $(PQ^T)_{(r,j)} = (\sum_{k,\, k \neq s} P_{(r,k)} * Q_{(j,k)}) + x * Q_{(j,s)}$
  - err(P,Q) becomes equation with one unknown
    - look at only terms involving x
    - get sum over j for which $R_{(r,j)}$ has a value of:

$$(R_{(r,j)} - (PQ^T)_{(r,j)})^2 = (R_{(r,j)} - (\sum_{k,\, k \neq s} P_{(r,k)} * Q_{(j,k)}) - x * Q_{(j,s)})^2$$

  - take derivative wrt x, set to 0, solve

6

Update step

Solution:

$$x = \frac{\sum_{j \text{ in } K} Q_{(j,s)} (R_{(r,j)} - \sum_{k \neq s} P_{(r,k)} * Q_{(j,k)})}{\sum_j Q^2_{(j,s)}}$$

for K the set of (r,j) for which $R_{(r,j)}$ has a value

- Similar equation if set element of Q to unknown y
- Iterate through elements of P, Q, repeatedly
- Find local minimum
  - improvement threshhold
- Need initial values P, Q

7

Stochastic gradient descent

- Define error $e_{ui} = R(u,i) - \sum_{\ell=1}^{f} (P(u,\ell) * Q^T(\ell,i))$
  - for each R(u,i) that has a value
  - residual factors
- Modification step:
  for u = 1 to M and i = 1 to N
    for ℓ=1 to f: $P(u,\ell)_{new} = P(u,\ell) + \gamma(e_{ui} * Q(i,\ell) - \lambda * P(u,\ell))$
    for ℓ=1 to f: $Q(i,\ell)_{new} = Q(i,\ell) + \gamma(e_{ui} * P(u,\ell) - \lambda * Q(i,\ell))$
- Choose γ and λ
- Repeat until error small enough
- Need starting values

8

Initializing

- Many strategies for initializing
- Example:
  - fill in each hole with average of column (item) values
  - decompose using SVD to get a rank f approximation ($U'_f \Sigma'_f V'^T_f$)
  - let $P_{init} = U'_f (\Sigma'_f)^{1/2}$
  - let $Q_{init} = V'_f (\Sigma'_f)^{1/2}$
  - note this particular initialization eliminates holes

9

Matrix factorization: summary

- Very effective method
- Issues:
  - Iteration is costly
    - Wait for local optimum?
  - Must choose initial values
- Subject of ongoing research

10

High-level issues for Collaborative Filtering:

Global effects

Effects over many or all of ratings

- ✓ different users have different rating scales
- metadata (attributes) for items and/or users
  - hybrid content/collaborative
- date of rating
- trend of user's ratings over time
- trend of item's ratings over time

Reference: Scalable Collaborative Filtering w/ Jointly Derived Neighborhood Interpolation Weights, Bell and Koren, *IEEE Intern. Conf. Data Mining* (part of winning Netflix contest team)

11

Final thought

All techniques we've seen behavior or topic oriented

What about links? What about PageRank?

12

Refining PageRank

$$\boldsymbol{pr} = (\alpha/n, \alpha/n, \ldots \alpha/n)^T + (1-\alpha)\, L^T \boldsymbol{pr}$$

- let $\boldsymbol{v} = (1/n, 1/n, \ldots 1/n)$
- rewrite $\boldsymbol{pr} = (\alpha)\boldsymbol{v}^T + (1-\alpha)\, L^T \boldsymbol{pr}$
- Refinement choices
  - change ***v***
  - change ***L***

"Topic Sensitive" PageRank

Haveliwala

- Use pre-defined topics
  - Open Directory Project (DMOZ)
    - "the largest, most comprehensive human-edited directory of the Web."
    - 16 top-level topics
- Each page has PageRank for each topic
  - Degree to which page is part of topic
- Calculate similarity of query to each topic
  - Use linear combination of topic PageRanks based on similarity values query to topic

Creating PageRank for a Topic

- set T_j contains all the URLS for j^{th} topic
- change PageRank equation:

$$pr_{new}(k) = \alpha(\mathbf{1/n}) + (1-\alpha)\sum_{i \text{ with edge from } i \text{ to } k} (pr(i) / t_i)$$

$$\downarrow$$

$$pr_{new}(k) = \alpha(\mathbf{t_k}) + (1-\alpha)\sum_{i \text{ with edge from } i \text{ to } k} (pr(i) / t_i)$$

where $t_k = 1/|T_j|$ if the k^{th} node represents a URL in T_j
0 otherwise

- removes random jumps to nodes outside the topic

Personalized PageRank

Kamvar et. al.

- Random leaps are biased by personal interests – change ***v***
- Combined with use of block structure to make more efficient:
  - Divide Web graph into blocks (clusters)
    - Use high-level domains (e.g. princeton.edu)
  - Calc. local PageRank within each block
  - Collapse each block into 1 node – new graph
    - Weighted edges between nodes
  - Calc. PageRank with biased leaps for block structure
  - Weight local PageRanks with block PageRank
    - Use to initialize power calculation

Refinement & Personalization Summary

- Looked at several techniques to modify search
- explicit user feedback
- user behavior: history
  - user history
  - crowd history
  - collaborative history: "people like you"
- role of social networks
  - general analysis
  - relationships
- models of recommender systems

17