Last time: Overview of techniques for refinement and personalization of searches

- query modification or ranking modification
- resources
  - user feedback
  - user history
  - crowd history
  - collaborative history: "people like you"

social networks
- friends etc
- general analysis

recommender systems
- content filtering
- collaborative filtering

1

Recommender Systems

2

Recommender Systems

- Look at classic model and techniques
  - Items
  - Users
  - Recommend Items to Users
- Recommend new items based on:
  - similarity to items user liked in past: individual history
    "Content Filtering"
  - Liked by other users similar to this user: collaborative history
    "Collaborative Filtering"
  - Liked by other users: crowd history
    - easier case

3

Recommender System attributes

- Need explicit or implicit ratings by user
  - Purchase is 0/1 rating
    - Movie tickets
    - Books
- Have focused category
  - examples: music, courses, restaurants
  - hard to cross categories with content-based
  - easier to cross categories with collaborative-based
    - users share tastes across categories?

4

Content Filtering

- Items must have characteristics
- user values item
  ⇒ values characteristics of item
- model each item as vector of weights of characteristics
  – much like vector-based IR
- user can give explicit preferences for certain characteristics

5

Content-based example

- user bought book 1 and book 2
  - what if actually rated?
- Average books bought = (0, 1, 0.5, 0)
- Score new books
  - dot product gives: score(A) = 0.5; score (B)= 1
- decide threshold for recommendation

	1st person	romance	mystery	sci-fi
book 1	0	1	1	0
book 2	0	1	0	0
new book A	1	.5	0	0
new book B	0	1	0	.2

6

Example with explicit user preferences

How use scores of books bought?

Try: preference vector p where component k =

user pref for characteristic k if ≠ 0

avg. comp. k of books bought when user pref =0

0 pref for user = "don't care"

p=(0, 1, 0.5, -5)

New scores?

p•A = 0.5

p•B = 0

	1st per	rom	mys	sci-fi
user pref	**0**	**1**	**0**	**-5**
book 1	0	1	1	0
book 2	0	1	0	0
new A	1	.5	0	0
new B	0	1	0	.2

Content Filtering: issues

- Vector-based one alternative
- Major alternatives based on machine-learning
- For vector based
  - how build a preference vector
    - how combined vectors for items rated by user
      - our example only 0/1 rating
    - how include explicit user preferences
  - what metric use for similarity between new items and preference vector
  - normalization
  - threshold?

Limitations of Content Filtering

- Can only recommend items similar to those user rated highly
- New users
  - Insufficient number of rated items
- Only consider features explicitly associated with items
  - Do not include attributes of user

Applying concepts to search

- Individual histories
  - Characterize individual by topic interest
    - Properties of objects interact with
  - Characterize query by related topics
    - Role of terms of query in topic
  - Modify ranking to prefer shared topics
  - Modify query to bias to shared topics

Example study: Personalizing Web Search Using Long-term Browsing History (in *WSDM11*)

- Goal: rerank
  - top 50 results from Google query
- Strategy:
  - score snippets from search result against user profile
  - rerank based on snippet score
- Selection of info for user profile
  - list of visited URLs w/ number visits
  - list of past search queries and pages clicked
  - list of terms with weights for content of pages visited

Personalizing Web Search Using Long-term Browsing History, cont

Studies selection of methods for

- user profile: what sources of terms use
- user profile: weights for terms
  - tf-idf
    - where get idf?
  - "modified BM25"- a "log odds measure"
- scoring
  - language model with adjustments for
    - URLs previously visited
    - original rank of snippet in search

worked best

performed best

$W_{modBM25}$ weighting

N = # documents on Web – estimated
n_{ti} = # docs on Web containing term t_i - estimated
R = # documents in user browser history
r_{ti} = # docs in user browser history that contain term t_i

$W_{modBM25}(t_i)$ =

$$\log \left(\frac{(\,(r_{ti} + 0.5)(N - n_{ti} + 0.5)\,)}{(\,(n_{ti} + 0.5)(R - r_{ti} + 0.5)\,)} \right)$$

13

Scoring

N_{si}= # unique words in snippet s_i
r_{si} = rank of snippet s_i in original search results
n_i = # previous visits by user to web page with snippet s_i
$w(t_k)$ = weigth of term t_k in user profile
w_{total} = sum of all term weights in user profile

$$\text{score}_{\text{lang. model}}(s_i) = \Sigma_{k=0}^{N_{si}} \log(\,(w(t_k) + 1)/w_{total})$$

- modif. for URLs previously visited:

$\text{score}_{w/URL}(s_i) = \text{score}(s_i)*(1+v*n_i)$ *parameter* v

- modif to acct. for orig. rank:

$\text{score}_{w/orig}(s_i) = \text{score}(s_i)*(\,1/(1+\log(r_{si}))\,)$

14

Personalizing Web Search Using Long-term Browsing History

Evaluation

- “offline” evaluation:
  - relevance judgments by volunteers
  - used to select best of algorithmic variations
- online evaluation of best variations:
  - add-on to Browser by volunteers
  - interleave original results (no personalization) with results reranked by snippet score
  - record clicks by user – which list from

15

Personalizing Web Search Using Long-term Browsing History

Results

- Offline: normalized DCG, avg. of 72 queries
  - Google’s ranking w/out personalization: 0.502
  - best-performing of variations for reranking: 0.573
- Online
  - 8% queries: # clicks from original and reranked same
  - of rest: 60.5% queries: more clicks from reranked
    39.5% queries: more clicks from original

Observation

- Reranking can be done completely in browser if enough space for data for user profile

16

What we’ve just seen:

Recommender systems: Content Filtering
Applying content filtering to search

Now back to recommender systems:

Collaborative Filtering

17

Collaborative Filtering

- Recommend new items liked by other users similar to this user
- need items already rated by user ***and other users***
- don’t need characteristics of items
  - each rating by individual user becomes characteristic
- Can combine with item characteristics
  - hybrid content/collaborative

18

Method types

- Nearest neighbor
  - Use similarity function
  - Prediction based on previously rated items
- Matrix Factorization
  - "Latent factors"
  - Matrix decomposition
- Both use (user × item) matrix
  - vector similarity

19

Example of nearest neighbor: Preliminaries

- Notation
  - $r(u,i)$ = rating of i^{th} item by user u
  - I_u = set of items rated by user u
  - $I_{u,v}$ = set of items rated by both users u and v
  - $U_{i,j}$ = set of users that rated items i and j
- Adjust scales for user differences
  - Use average rating by user u:

    $r_u^{avg} = (1/|I_u|) * \sum_{i \text{ in } I_u} r(u,i)$
  - Adjusted ratings: $r_{adj}(u,i) = r(u,i) - r_u^{avg}$

20

One choice of similarity function: User Similarities

- similarity between users u and v
  - Pearson correlation coefficient

$$sim(u,v) = \frac{\sum_{i \text{ in } I_{u,v}} (r_{adj}(u,i) * r_{adj}(v,i))}{\left(\sum_{i \text{ in } I_{u,v}} (r_{adj}(u,i))^2 * \sum_{i \text{ in } I_{u,v}} (r_{adj}(v,i))^2 \right)^{1/2}}$$

21

Predicting User's rating of new item: User-based

For item i not rated by user u

$$r^{pred}(u,i) = r_u^{avg} + \frac{\sum_{v \text{ in } S} (sim(u,v) * r_{adj}(v,i))}{\sum_{v \text{ in } S} |sim(u,v)|}$$

S can be all users who have rated i or just those users *most similar* to u

22

Collaborative filtering example

user ratings

	book 1	book 2	book 3	book 4
user 1	5	1	2	0
user 2	x	5	2	5
user 3	3	1	x	2
user 4	4	0	2	?

adj. user ratings

	book 1	book 2	book 3	book 4
user 1	3	-1	0	-2
user 2	x	1	-2	1
user 3	1	-1	x	0
user 4	2	-2	0	?

Collaborative filtering example

- $sim(u1,u4) = (6+2)/(10*8)^{1/2} = .894$
- $sim(u2,u4) = (-2)/(5*4)^{1/2} = -.447$
- $sim(u3,u4) = (2+2)/(2*8)^{1/2} = 1$
- predict $r(u4, book4) = 2 + \frac{(-2)*.894 + 1*(-.447) + 0*1}{.894 + .447 + 1}$

  $= 2 - .955 \approx 1$

24

Another choice of similarity function: Item Similarities

- similarity between items i and j
  - vector of ratings of users in $U_{i,j}$
  - cosine measure using adjusted ratings

$$\text{sim}(i,j) = \frac{\sum_{u \text{ in } U_{i,j}} (r_{adj}(u,i) * r_{adj}(u,j))}{\left(\sum_{u \text{ in } U_{i,j}} (r_{adj}(u,i))^2 \ \sum_{u \text{ in } U_{i,j}} (r_{adj}(u,j))^2 \right)^{1/2}}$$

25

Predicting User's rating of new item: Item-based

For item i not rated by user u

$$r^{\text{item-pred}}(u,i) = \frac{\sum_{j \text{ in } T} (\text{sim}(i,j) * r(u,j))}{\sum_{j \text{ in } T} |\text{sim}(i,j)|}$$

T can be all items in I_u or just items *most similar* to i

- Prediction uses only u's ratings, but similarity uses other users' ratings

26

Limitations

- May not have enough ratings for new users
- New items may not be rated by enough users
- Need "critical mass" of users
  - All similarities based on user ratings

But can take user "out of comfort zone"

27

Applying nearest-neighbor collab. filtering concepts to search

- Collaborative histories
  - How determine user similarity?
    - Behavior on identical searches?
    - Overlap of general topic interests?
      - From overlapping behaviors
      - Hybrid content-based and behavior-based
    - Computational expense?
      - Argues for general topic-interest characterizations
  - How apply similarity?
    - Same search? Bias ranking?
    - Same topic of search? Bias topics of results?

28

Example

from A Large-scale Evaluation and Analysis of Personalize Search Strategies (in *WWW07*)

- Goal: rerank search results
- Based on query log history – clicks as ratings
- Also uses 67 pre-defined topic categories
- Strategy:
  - get similarity of users based on user history of visited pages
  - find K most similar users to user doing search
    K nearest neighbor; use K=50
  - calc. score for each result of search based on click history of K nearest neighbors
  - rerank results of search based on score

29

Details

from A Large-scale Evaluation and Analysis of Personalize Search Strategies (in *WWW07*)

P(u) = collection of Web pages visited by user u in the past

$$P(p|u) = \frac{\#\text{ times u clicked on page p in past}}{\text{total \# time u clicked on a page in past}}$$

w(p) = log(total # users / # users visited page p)
"impact weight" - idf-like

c(p) = "category vector" for page p
do classification of page
vector gives confidence # for top 6 categories (other entries 0)

User profile $\quad \mathbf{c}_\ell(u) = \Sigma_{p \text{ in } P(u)} P(p|u)w(p)\mathbf{c}(p)$

User similarity $\quad \text{sim}(u_1, u_2) = \dfrac{\mathbf{c}_\ell(u_1) \bullet \mathbf{c}_\ell(u_2)}{||\mathbf{c}_\ell(u_1)|| \ ||\mathbf{c}_\ell(u_2)||}$

30

Details

from A Large-scale Evaluation and Analysis of Personalize Search Strategies (in *WWW07*)

$S_k(u_a)$ denotes k nearest neighbors of user u_a

click history:

$|clicks(q,p,u_s)|$ = # clicks on pg p by user u_s on past query q

$|clicks(q,*,u_s)|$ = # clicks overall by user u_s on past query q

the score of a page p for query q and user u:

$$S(q,p,u) = \frac{\Sigma_{u_s \text{ in } S_k(u)} \; sim(u_s,u) * |clicks(q,p,u_s)|}{\beta + \Sigma_{u_s \text{ in } S_k(u)} \; |clicks(q,*,u_s)|}$$

β is a “smoothing factor”; taken to be 0.5

31

Experiments

from A Large-scale Evaluation and Analysis of Personalize Search Strategies (in *WWW07*)

- Data set: MSN query logs 12 days August 2006
  sampled 10,000 distinct users
  used 11 days for training, last day for testing
  ~ 4000 test queries
- Action, for each user and query
  – re-rank top 50 results using a “fusion” of original rank and order given by page scores S(q,p,u)
- Evaluation: 2 metrics
  1. a DCG-like metric with clicking indicating relevance
  2. average rank of clicked items

32

Results

from A Large-scale Evaluation and Analysis of Personalize Search Strategies (in *WWW07*)

- Good news:
  re-ranking improves over original ranking
- So-so news:
  improvement is 3.62% on queries where there is room for improvement
- Not so good news:
  non-collaborative personalization improves 3.68%

$$S(q,p,u) = \frac{|clicks(q,p,u)|}{\beta + |clicks(q,*,u)|}$$

33