Refining and Personalizing Searches

1

Targets

- collection
- ➢ query
- satisfying documents
  - increase set?
- ➢ ranking

2

Themes

- Explicit feedback versus search history
- Personalized history versus group history

3

Refine initially: query

- Help user get better query
- Commonly, query expansion
  - add synonyms
    - Improve recall
    - Hurt precision?
    - Sometimes done automatically – with care
  - Modify based on prior searches
    - Not automatic
    - All prior searches - eg. suggested search terms
      vs
    - *your* prior searches

4

Refining after search

- Use user feedback
  or
- Approximate feedback with first results
  - Pseudo-feedback
  - Example: “Yahoo assist” (?still)
- change ranking of current results
  or
- search again with modified query

5

Explicit user feedback

- User must participate
- User marks (some) relevant results
  or
- User changes order of results
  - Can be more nuanced than relevant or not
  - Can be less accurate than relevant or not
    - Example: User moves 10th item to first
      - says 10th better than first 9
      - Does not say which, if any, of first 9 relevant

6

User feedback in classic vector model

- User marks top p documents for relevance
  - p = 10 to 20 “typical”
- Construct new weights for terms in query vector
  - Modifies query
  - Could use just on initial results to re-rank

7

Deriving new query for vector model

For collection C of n doc.s

- Let C_r denote set all relevant docs in collection,

Perfect knowledge Goal:

Vector $\mathbf{q}_{opt}$ =

$1/|C_r|$ * (sum of all vectors d_j in C_r) -

$1/(n- |C_r|)$ * (sum of all vectors d_k not in C_r)

centroids

8

Deriving new query for vector model: Rocchio algorithm

Give query **q** and relevance judgments for a subset of retrieved docs

- Let D_r denote set of docs judged relevant
- Let D_{nr} denote set of docs judged not relevant

Modified query:

Vector $\mathbf{q}_{new}$ = $\alpha\mathbf{q}$ +

$\beta/|D_r|$ * (sum of all vectors $\mathbf{d}_j$ in D_r) -

$\gamma/(|D_{nr}|)$ * (sum of all vectors $\mathbf{d}_k$ in D_{nr})

For tunable weights α, β, γ

9

Remarks on new query

- α: importance original query
- β: importance effect of terms in relevant docs
- γ: importance effect of terms in docs not relevant

- Usually terms of docs not relevant are least important
  - Reasonable values α=1, β=.75, γ=.15
- Reweighting terms leads to long queries
  - ***Many*** more non-zero elements in query vector $\mathbf{q}_{new}$
  - Can reweight only most important (frequent?) terms
- Most useful to improve recall
- Users don’t like: work + wait for new results

10

Simple example user feedback in vector model

- **q** = (1,1,0,0)
- Relevant: **d1** = (1,0,1,1)
  d2 = (1,1,1,1)
- Not relevant: **d3**=(0,1,1,0)
- α, β, γ = 1
- $\mathbf{q}_{new}$ = (1,1,0,0) + (1, 1/2, 1, 1) - (0,1,1,0)
  = (2, 1/2, 0, 1)

Term weights change New term

Observe: Can get negative weights

11

Re-ranking using explicit feedback

- Algorithms usually based on machine learning
  - Learn ranking function that best matches partial ranking given

- Simple example
  - 2007ish: Google experiment; only affects repeat of same search
  - 2008: became SearchWiki feature for Google accounts
  - 2010: functionality reduced to “starred” results list
  - 2012: replaced by +1?

12

Implicit user feedback

- Click-throughs
  - Use as relevance judgment
  - Use as reranking:
    When click result, moves it ahead of all results didn't click that come before it
  - Problems?
- Better implicit feedback signals?

13

Behavior History

- Going beyond behavior on ***same*** query.
- Personal history versus Group history
- Group history
  - Primarily search history
    - Google's claim Bing copies
- Personal history
  - Searches
  - Other behavior – browsing, mail?, …
  - Characterize interests: topics

14

Collaborative history

Group history + personal history =>
History of people "like" you

How characterize?
- Shared behaviors
- Shared topics

15

Example: 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-based"
  - Liked by other users similar to this user: collaborative history
    "Collaborative Filtering"
  - Liked by other users: group history
    - easier case

16

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?

17

Content-based recommendation

- 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

18

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

19

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

20

Content-based: 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?

21

Limitations of Content-based

- 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

22

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 query to bias to shared topics
  - Modify ranking to prefer shared topics

23

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

24

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

Equations for Pers'lizing Web Search Using Long-term Browsing History

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

$W_{modBM25}(t_i) = \log(((r_{ti} + 0.5)(N-n_{ti}+0.5)) / ((n_{ti} + 0.5)(R - r_{ti} + 0.5)))$

N_{si}= # unique words in snippet i
r_{si} = rank of snippet 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

$score_{lang.\ model}(s_i)$ = sum over k=0 to N_{si} of $\log ((w(t_k) +1)/w_{total})$

modif. for URLs previously visited: $score_{w/URL}(s_i)= score(s_i)(1+v*n_i)$
where v is a parameter; v=10 is used in the experiments

modif to acct. for orig. rank: $score_{w/orig}(s_i) = score(s_i)(1/(1+\log (r_{si})))$

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

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
  - 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

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

Method types

(see Adomavicius and Tuzhilin paper)

- Memory-Based
  - Similar to vector model
  - Use (user × item) matrix
  - Use similarity function
  - Prediction based on previously rated items
- Model-Based
  - Machine-learning methods
  - Model of probabilities of (users × items)

Memory-Based: 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: $r_{adj}(u,i) = r(u,i) - r_u^{avg}$

31

One Memory-Based method: User Similarities

- similarity between users u and v
  - Pearson correlation coefficient

$$sim(u,v) = \frac{\sum_{i \text{ in } I_{u,v}} (r_{adj}(u,i) * r_{adj}(v,i))}{\left(\sum_{i \text{ in } I_{u,v}} (r_{adj}(u,i))^2 * \sum_{i \text{ in } I_{u,v}} (r_{adj}(v,i))^2 \right)^{½}}$$

32

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 or just users *most similar* to u

33

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 $r(u4, book4) = 2 + \dfrac{(-2)*.894 + 1*(-.447) + 0*1}{.894 + .447 + 1}$

  $= 2 - .955 \approx 1$

35

One Memory-Based Method: Item Similarities

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

$$sim(i,j) = \frac{\sum_{u \text{ in } U_{i,j}} (r_{adj}(u,i) * r_{adj}(u,j))}{\left(\sum_{u \text{ in } U_{i,j}} (r_{adj}(u,i))^2 \ \sum_{u \text{ in } U_{i,j}} (r_{adj}(u,j))^2 \right)^{½}}$$

36

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 or just items *most similar* to i

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

37

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)

38

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

39

Applying 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?

40

Example

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

- Goal: rerank search results
- Based on query log history – clicks
- 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

41

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

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

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

42

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:

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

β is a "smoothing factor"; taken to be 0.5

43

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 $\boldsymbol{v}$
  - change $\boldsymbol{L}$

44

"Topic Sensitive" PageRank

Haveliwala

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

45

Personalized PageRank

Kamvar et. al.

- Random leaps are biased by personal interests – change $\boldsymbol{v}$
- 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

46

Summary

- Looked at several techniques for modifying search
  - Explicit User feedback
    - revise query
  - Implicit User feedback – behavior history
    - Individual history
    - Group history
    - Collaborative history
  - Recommender systems
  - Modifying PageRank

47