Enrico Minack, Gianluca Demartini, Wolfgang Nejdl

CURRENT APPROACHES TO SEARCH RESULT DIVERSIFICATION

Motivation

The Web grows, the number of relevant results grows as well

Search engine users look only at top few documents

They should be a *good* sample of the entire relevant set

Outline

Diversification of Web results: problem definition

A Framework for search result diversification

- Relevance Functions
- Similarity measures
- Objective functions

Datasets and evalution techniques

Room for improvement

Diversifying Web Search Results

Levels of diversity in Web Sarch

- Ambiguous queries: different senses
- Clear queries: different aspects/subtopics

Problem: find the subset with the *k* most **relevant** and **diverse** results

In a ranked list:

- Top-k docs results are relevant
- i-th result should be novel compared to the i-1 previous docs

Diversifying Web Search Results

Different types of diversity exists

- Topic, Opinion, Genre, Document type, Time, Conflicting info, ...

Different applications can benefit from *result diversification*

- Web Search
- News
- Blogs
- Product Search
- ...

Trade-off relevance/novelty

Finding the optimal set of items which is both relevant and diverse

- Relevance measure
- Similarity (diversity) measure
- Diversification objective (trade-off)
  - NP-hard problem
  - Use greedy algorithms
  - Compute an aproximation

Relevance measure

All systems work on top-k items ordered by a relevance measure

For both full text and structured datasets

Different measures can be used to identify such set:

- Language models [1]
- Vector space [2]
- KL-divergence [4]
- ...

Similarity measures

- Semantic Distance (Textual similarity)
  - Cosine sim
  - Jaccard sim [1]
  - Euclidean distance [2]
- Categorical distance
  - Tree distance based on taxonomies [1] [3]
  - Order of attributes to be diversified [5]
- Novel Information
  - KL Divergence [2]

- No measure exploits genre, sentiment, or other diversity types

Objective functions

Combining relevance and diversity

Find the optimal set of items which is relevant and diverse

Proposed objective functions:

- Max-sum [1]: weighted sum
- Max-min [1]: min relevance and dissimilarity
- Average dissimilarity [1]: adds to the relevance the avg dissimilarity
- Max-sum of max-score [5]: max diveristy after max relevance
- Max-product [4]: select i-th results by relevance*dissimilarity(i,1..i-1)
- Categorical diversification [3]: covered categories

The problem is NP-hard

Aproximations use **on-line** greedy algorithms

Datasets

Main distiction is between full text vs structured datasets

Full text:

- Top k docs from commercial search engines [3]
- TREC Interactive [4]

Structured data:

- Yahoo! Autos [5]
- DB [2]
- Syntetic datasets [2]

Ground truth:

- Wikipedia disambiguation pages [1]
- Amazon Mturk [3]

Evaluation Measures

New diversity aware measures are defined for IR tasks only

- alpha-NDCG [6]: relevance based on subtopics covered in the query and contained in previous results
- S-Precision, S-Recall (aka novelty [1]), WS-Precision [4]
- NDCG-IA MAP-IA MRR-IA [3]: user intent

DB search

- goodness of the aproximation compared to the optimal result
- efficiency

TREC 2009 Web Track

Diversity Task

- Return a ranked list that provides complete coverage for a query
- Avoiding redundancy in the result list

Subtopics, each related to a different user need

- For each subtopic, assessors make a binary relevance judgment

Measures:

- α-nDCG
- MAP-IA
- give no credit to duplicate and near-duplicate documents

http://plg.uwaterloo.ca/~trecweb/

Example topic

Topic: physical therapist

Subtopics (not given!):

- What does a physical therapist do?
- Where can I find a physical therapist?
- Therapy cost per hour
- Required Training
- American Physical Therapy Association
- Salary
- Difference between a occupational therapist and a physical therapist
- Required education

Topical diversity

Possible next steps

Algorithms

- Off-line steps to simplify the on-line optimization step
  - Relevance functions focusing on diversity (no re-ranking)
    - [5] proofs that inverted indexes can not do that
- Other diversity notions: similarity measures not based on content
  - Opinion, topic, genre, time, ...
  - Combine different notions in one measure

Interaction

- What diversity the user expects?

Benchmarks

- TREC is producing a topical-diversity benchmark
- One corpus for each notion of diversity should be created

References

[1] Gollapudi, S., Sharma, A.: An Axiomatic Approach for Result Diversification. In: WWW '09

[2] Jain, A., Sarda, P., Haritsa, J.R.: Providing Diversity in K-Nearest Neighbor Query Results. In: PAKDD04

[3] Agrawal, R., Gollapudi, S., Halverson, A., Ieong, S.: Diversifying Search Results. In: WSDM09

[4] Zhai, C.X., Cohen, W.W., Laerty, J.: Beyond Independent Relevance: Methods and Evaluation Metrics for Subtopic Retrieval. In: SIGIR03

[5] Vee, E., Srivastava, U., Shanmugasundaram, J., Bhat, P., Yahia, S.A.: Efficient Computation of Diverse Query Results. In: ICDE08

[6] Clarke, C.L., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Buettcher, S., MacKinnon, I.: Novelty and Diversity in Information Retrieval Evaluation. In: SIGIR08