Overview of Personalized Information Retrieval Track @ FIRE-2011

Debasis Ganguly, Johannes Leveling, Keith Curtis, Wei Li, Gareth Jones
School of Computing, CNGL, Dublin City University
Outline

- TREC-style evaluation
- Beyond TREC: Personalized IR task objectives
- Data preparation
- Information retrieval with logs
- Results
- Conclusions and future work
TREC-style Evaluation

- Information Retrieval evaluation campaigns
  - TREC, CLEF, NTCIR, INEX, FIRE

- Evaluation methodology:
  - Organizer(s):
    - create topics (queries)
  - Participating systems:
    - index document collection, process topics, submit results
  - Organiser(s) (+ participants):
    - evaluate submissions
Example TREC topic

<num>401</num>
<title> foreign minorities, Germany </title>
<desc> What language and cultural differences impede the integration of foreign minorities in Germany? </desc>
<narr> A relevant document will focus on the causes of the lack of integration in a significant way; that is, the mere mention of immigration difficulties is not relevant. Documents that discuss immigration problems unrelated to Germany are also not relevant. </narr>
PIR Task Motivation (beyond TREC)

- Limitations of ad-hoc IR evaluation paradigm:
  - One topic (query) fits all (users)
  - One result set fits all (users)

- PIR task: Log the topic development to enable research in personalisation
  - Different users have different ways of expressing the same information need.
  - Different users formulate topics for same broad search categories (e.g. “Bollywood movies”)
  - Users are not really sure what they are looking for initially. So querying is an iterative process e.g. “education in india” -> “top engineering institutes of India” -> “research in IIT Bombay” etc.
PIR Task Motivation (contd.)

**Elements of personalization:**
- Different query formulations and relevant documents
- Same user develops topic and judges relevance
- Topic development and evaluation on same corpus
  \( \rightarrow \) reproducible results

**Elements of collaboration:**
- Users choose a search category as their starting points.
- Two users with same category indicate users with similar interests.

**Research Question:**
- Can we tune IR systems to address individual user-specific information needs?
Task Based Navigation

Select category

Form and execute a query

Read docs and reformulate query

Enter a final test query which will be assigned to you for relevance judgement

Indian paintings and painters

Indian painters

Kangra paintings

Hussain paintings

Hussain controversy paintings

Hussain controversial paintings

M.F. Hussain controversy paintings

Find a detailed information on M.F. Hussain's controversial paintings

Information about the reasons for controversy, Hussain's reactions to the controversies are relevant here. A third party critic's view over the matter is also relevant.
Difference with ad-hoc topics

**TREC:** $Q_1 \rightarrow D_1, \ldots, D_m$
(single query, single result set)

**PIR:** $(Q_{1k} \rightarrow D_{1k}, \ldots, D_{mk}) \times (n-1) \rightarrow Q_n$
(multiple users, multiple related queries, multiple related result sets)
PIR track activity flow

1. Category selection
2. Query formulation
3. View/browse result list
4. Topic summarization
5. Final topic formulation
6. Relevance assessment

Impact of strikes

User interface

Topic developer

Category list
Indexed documents
Search logs
TREC-style topics
Retrieval runs pool
qrels

Participating IR systems
Evaluation
Data Preparation (1/3)

- **Document collection** –
  - English FIRE-2011 ad-hoc collection (articles from Indian and Bangladesh newspapers)
  - Index the collection with Lucene
  - Identify 15 broad category news domains

- **Java Servlet based search interface** which supports
  - user registration to maintain user identity
  - category selection and navigation
  - document retrieval against a user query
  - viewing and bookmarking documents
  - submission of summary and final test topic.
Data Preparation (2/3)

Each line in a CSV formatted log contains:

- User name (U), category name (C) – to identify the current session.
- Query string (Q) – to identify the current query.
- Document ID (D) – to identify the current document on which an action is performed
- Action – click to view or bookmark
- Timestamp – to compute relative viewing times
Data Preparation (3/3)

- Queries in extended TREC format (final test topics)
- Additional fields:
  - User name of the topic developer
  - Topic category
Information Retrieval using Logs

Question:
- Can we tune IR systems to address individual user-specific information needs?

Objective:
- Investigate benefit of using additional information about a user (topic developer) on IR performance

Data:
- Ad hoc document collection
- User information (search history+search categories)
- Final topics

Evaluation metrics:
- P@5, P@10, (MAP)
Not everything goes according to plan 😞

- 26 registered participants, but no run submissions!
- 25 topics: enough for standard IR evaluation, but not enough for PIR
- 10 topic developers with different interests
- Very small overlap
  - Categories:
    - Indian tourism – 5
    - Relation of India with its neighbouring countries – 4
    - Indian traditions and customs – 3
    - Indian paintings and painters – 3
  - Documents: Virtually no overlap – at most 2
We generated three baseline submissions:

- **BL1**: standard IR run on test topic titles (no information from logs used!)
  - Pseudo-relevance Feedback (PRF) with $R=10$ and $T=10$

- **BL2**: standard IR run using intermediate query titles of a given test topic as a query
  - PRF with $R=10$ and $T=10$

- **BL3**: standard IR run using viewed documents as additional pseudo-relevant documents
  - PRF with $R \geq 10$ and $T=10$

Document pool constructed from baseline runs
A data structure for efficient log processing

- Two level hash table with list of list
- Quickly retrieves the intermediate queries and viewed docs for a given category and user name
Results

A sample instance:

Topic number 1 “rock climbing india”

  BL1: Ranks of relevant docs = {27,184,303}
  BL2: Ranks of relevant docs = {2,3,4}
  BL3: Ranks of relevant docs = {1,2,4}
Conclusions and Future Work

- Results show that logs can be used to improve precision at top ranks.

- If two simple approaches work reasonably well, then more complex methods may work even better. For example:
  - Using the view times to predict relevance
  - Using the bookmarks as pseudo-relevant documents
  - Using RS techniques such as popularity of a document across users
Future Work

- Identify reasons for lack of submissions
- Generate more logs from more users
- Make the search task more interesting by using Wikipedia instead of news
Please provide your suggestions and comments and express your interest for a PIR task 2012

THANK YOU… 😊