Effective Relevance Feedback for Entity Ranking

Tereza Iofciu, **Gianluca Demartini**,
Nick Craswell, Arjen P. de Vries

Ranking People

- Expert Finding in TREC-ENT (Enterprise Track)
- Collection:
  - Corpus: crawl of *.w3.org sites
  - People: names of 1092 people who may be experts
- Query:
  - `'information retrieval'`
- Results:
  - A **list of people** who know about information retrieval

Ranking Actors

- Queries are lists of actors on the Web, e.g.
  - Query: 1930s
    - Answers: Fred Astaire, Charlie Chaplin, W.C. Fields, Errol Flynn, Clark Gable, Greta Garbo, etc
  - Query: action
    - Answers: Arnold Schwarzenegger, etc

Ranking…

- People
  - Expert Finding evaluation
- Actors
  - No evaluation initiative… yet?!
- Car companies, countries, museums, …

  [i.e., insert your fav entity type here]

Entity Ranking!!!

Example INEX XER 2008 Topics

- Countries that have hosted FIFA Football World Cup tournaments: *countries; football world cup*

- Formula 1 drivers that won the Monaco Grand Prix: *racecar drivers; formula one drivers*

- Italian Nobel prize winners: *nobel laureates*

…

Many examples on
http://www.ins.cwi.nl/projects/inex-xer/topics/

Entity Ranking

- Topical query Q
- Entity (result) type T_X
- A list of entity instances Xs

- Systems employ XML element text, structure, links

Topic 116

Q	**Title** Italian Nobel prize winners
Xs	**Entities** Dario Fo (#176791) Renato Dulbecco (#744909) Carlo Rubbia (#44932)
T_X	**Categories** Nobel laureates (#924)

Description
I want all the Italian people who won the Nobel prize.
Narrative
I want a list of people who were Nobel prize laureates in any field and have Italian nationality.

INEX XER Tasks

- Entity Ranking (ER)
  - Given Q and T, provide Xs
- List Completion (LC)
  - Given Q and Xs[1..m]
  - Return Xs[m+1..N]

INEX XER 2008 Assumptions

- Entities (Xs) are represented as Wikipedia pages
- Binary relevance, MAP (`xinfAP*`)

* A simple and efficient sampling method for estimating AP and NDCG.
  Emine Yilmaz, Evangelos Kanoulas, and Javed A. Aslam. SIGIR'08

AND NOW…

HOW TO SOLVE
ENTITY RANKING

Possible approaches to XER

- Link structure [Pehcevski et al. ECIR08]
- Language Models [Weerkamp et al. INEX08]
- NLP based [Demartini et al. LA-WEB08]
- Ontology based [Demartini et al. WISE08]
- Passage retrieval [Zaragoza et al. CIKM07]

- It is a recent task (2y): low effectiveness
- All previous work use categories

Wikipedia Category structure

- Category structure
  - Unweighted entity-category relation
  - Missing categories
  - Noise (categories not expressing type)

- Our contribution:
  - Find good categories and improve results via Relevance Feedback

Relevance feedback (RFB)

- User issues a query to a system
- Pseudo RFB
  - System uses top N retrieved entities to refine results
- Interactive RFB
  - User selects relevant entities in top K results
  - System uses relevant entities in top N to refine results

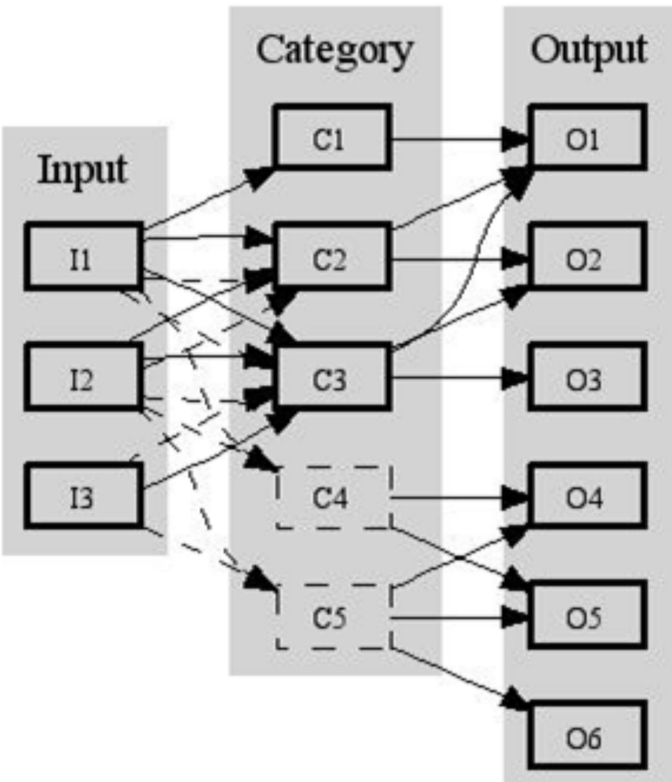
Relevance feedback for XER

- Propagation of weights through a DAG for finding best categories

- Edges between article *i* and category *j*
  - Hard edges: article *i* belongs to category *j*
  - Soft Edges: article *i* links to article *i′* in category *j*

Relevance feedback for XER

- h_j: total incoming hard edges for category j
- s_j: total incoming soft edges for category j



$$catweight_j = \frac{10^{h_j} + s_j}{\log(catsize + 50)}$$

$$entityweight_k = \left(\sum_{j=1}^{n} catweight_j\right) * P_k$$

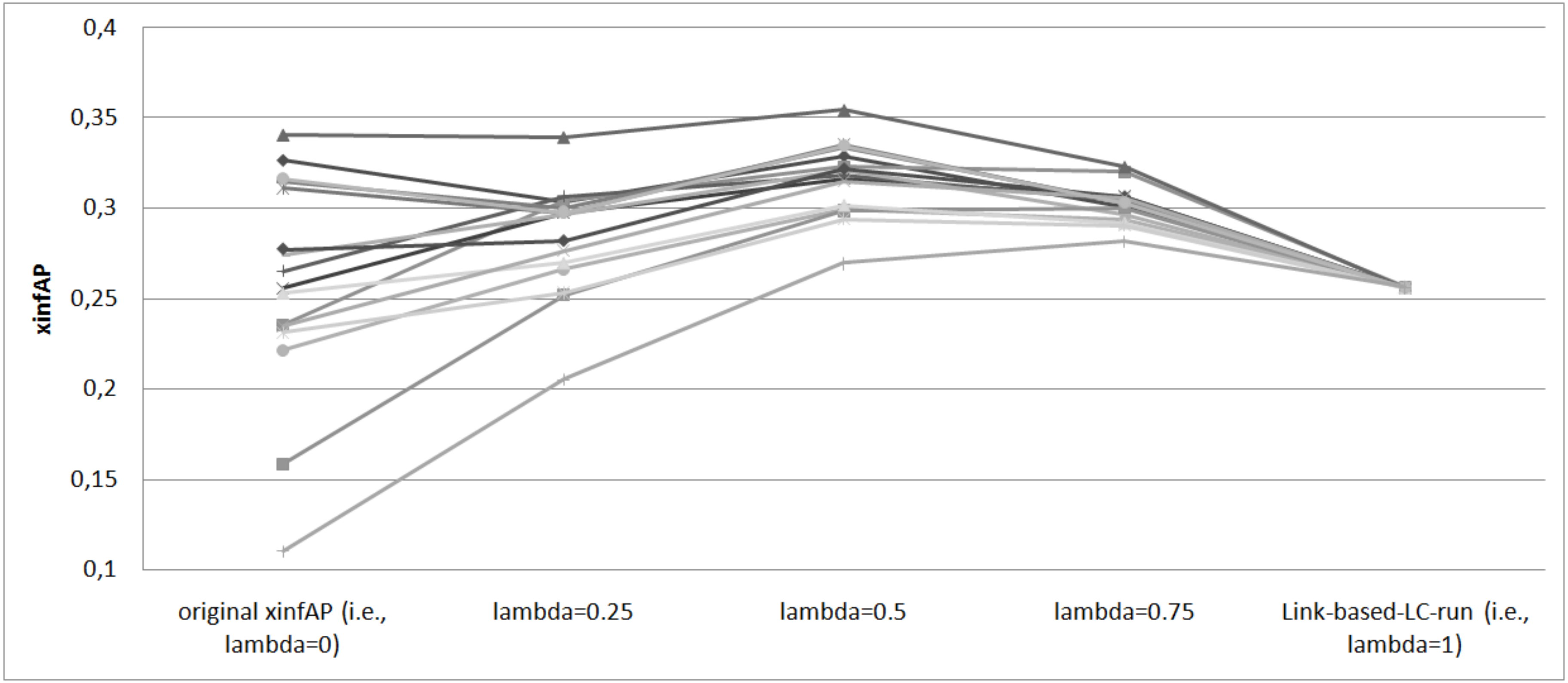
Setting

- INEX XER 2008
  - 600k Wikipedia articles
  - 35 topics
  - 32 Runs used as baselines
- Seed for our algorithm
  - Example entities
  - Pseudo RFB
  - Interactive RFB
- Fusion with baseline

Experimental Results

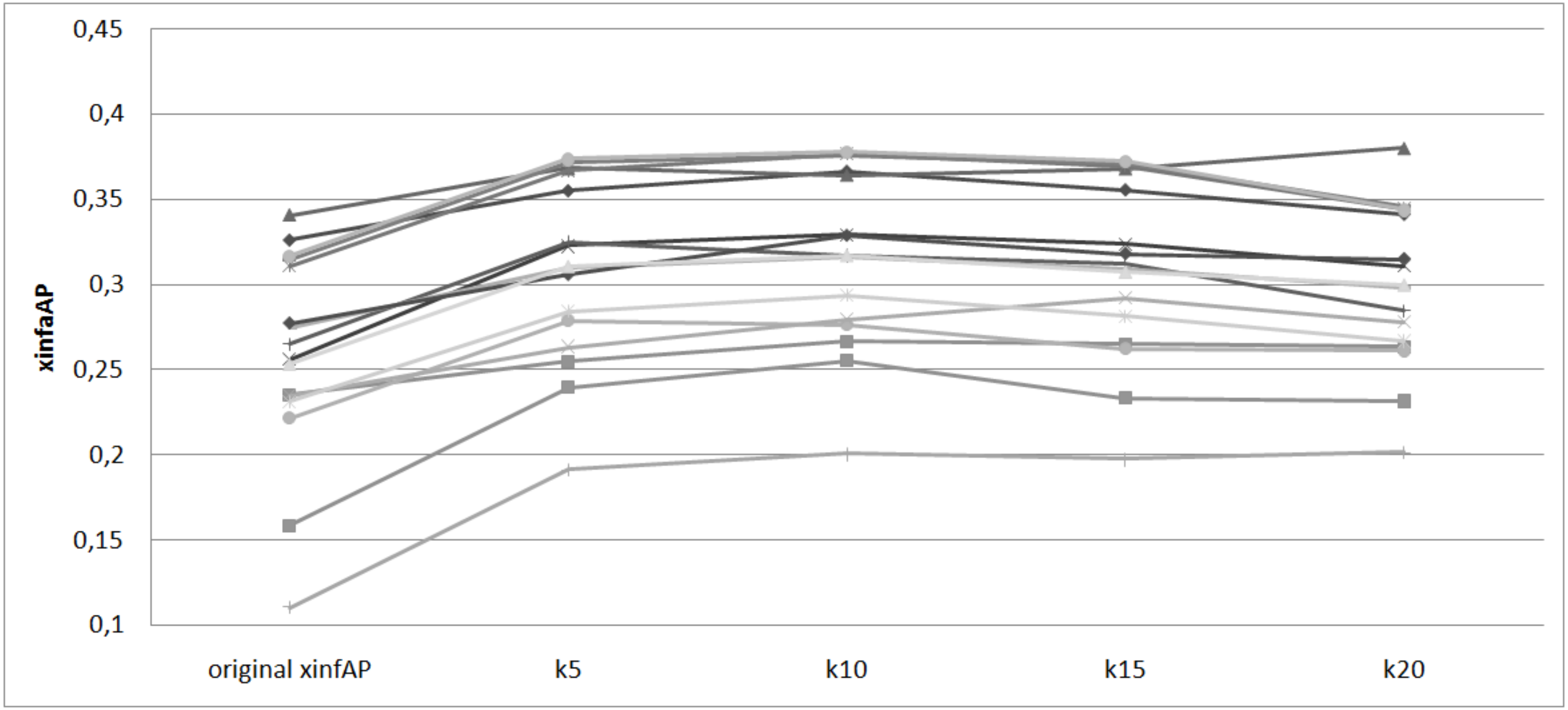
- Fusion with baseline (Example entities)

$$score(e,q) := \lambda \cdot baseline(e,q) + (1-\lambda) \cdot LinkBased(e,q)$$



Experimental Results

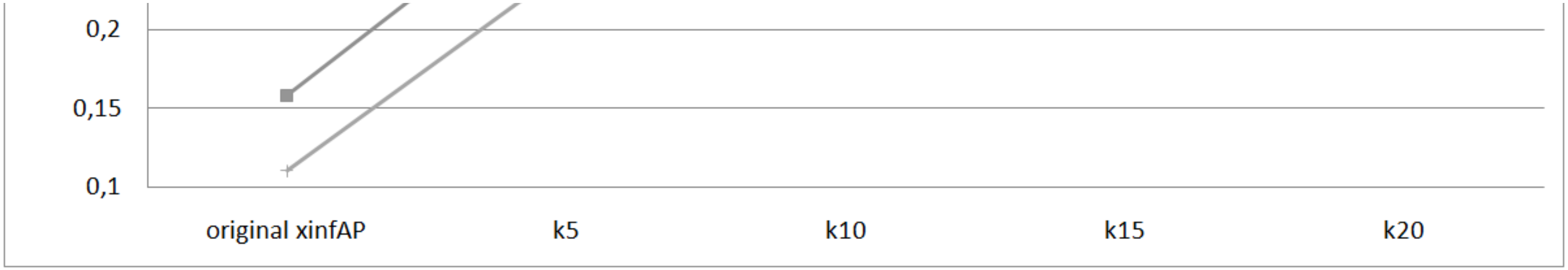
- Fusion with baseline (Pseudo RFB)



Experimental Results

- Table 5: Average absolute improvement of xinfAP for different values of k in the pseudo relevance feedback and in the interactive relevance feedback cases.

	K=5	K=10	K=15	K=20
Pseudo RFB	0.050	0.057	0.051	0.040
Interactive RFB	0.083	0.103	0.112	0.118
AVG relevant in top-k	1.92	3.57	4.88	6.04



Experimental Results

- Early Precision

Table 6: Expected P@20 measured for different values of k in the interactive relevance feedback case.

	K=5	K=10	K=15	K=20
baseline	0.204	0.204	0.204	0.204
combination $\lambda = 0.5$	0.222	0.225	0.228	0.247
Relevance Feedback	0.265	0.289	0.289	0.29

Unique contributions

- Pseudo RFB

	K=5	K=10	K=15	K=20
relevant in baseline	5.746	4.820	4.770	4.752
relevant in pseudo RF	3.932	4.357	4.393	4.227
relevant in both	12.146	13.073	13.123	13.141
missed relevant	4.489	4.064	4.029	4.195

- Interactive RFB

	K=5	K=10	K=15	K=20
relevant in baseline	8.191	6.545	5.941	5.470
relevant in pseudo RF	2.450	3.114	3.496	3.677
relevant in both	9.702	11.348	11.952	12.423
missed relevant	5.971	5.307	4.925	4.745

Per topic analysis

- RFB for doc search
  - works well on average
  - on some queries performs badly
- 15 topics over 12 systems had 0 relevant after RFB
  - All system having low AvgPrec: not a good seed
- 3 topics over 6 systems having 0 AvgPrecs had 0.16 after RFB
  - Exceptional case: 40 relevant

Conclusions

- Using top 10 results yields to best results
- Interactive RFB yields to best results
- RFB for XER is beneficial
  - differently from doc search
  - for all retrieval methods
- Limitations
  - Single test collection
  - Non-optimal paramenters for the model

Thanks