

Bloggers as Experts

Krisztian Balog
Maarten de Rijke
Wouter Weerkamp

ISLA, University of Amsterdam



task

find people that are experts on a given topic
people-topic association finding

models

two models to estimate the probability of a candidate expert (ca) given the query (q) [1]

$$p(ca|q) = \frac{p(q|ca)p(ca)}{p(q)}$$

model 1

$$\prod_{t \in q} \left[(1 - \lambda) \cdot \left(\sum_d p(t|d) \cdot p(d|ca) \right) + \lambda \cdot p(t) \right]^{n(t,q)}$$

model 2

$$\sum_d \prod_{t \in q} [(1 - \lambda) \cdot p(t|d) + \lambda \cdot p(t)]^{n(t,q)} \cdot p(d|ca)$$

results

expert finding task of TREC Enterprise track

model	2005			2006			2007		
	MAP	P@10	MRR	MAP	P@10	MRR	MAP	P@10	MRR
1	0.1883	0.3060	0.4692	0.3206	0.5510	0.7264	0.3700	0.1260	0.5303
2	0.2053	0.3000	0.6088	0.4660	0.7388	0.9354	0.4137	0.1620	0.5666

task

find blog(ger)s that have a principal, recurring interest in a given topic
blogger-topic association finding

models

we adjust the expert finding models to estimate the probability of a blog given the query (q)

$$p(blog|q) = \frac{p(q|blog) \cdot p(blog)}{p(q)}$$

blogger

$$\prod_{t \in q} \left[(1 - \lambda_{blog}) \cdot \left(\sum_{post \in blog} p(t|post, blog) \cdot p(post|blog) \right) + \lambda_{blog} \cdot p(t) \right]^{n(t,q)}$$

posting

$$\sum_{post \in blog} \prod_{t \in q} [(1 - \lambda_{post}) \cdot p(t|post) + \lambda_{post} \cdot p(t)]^{n(t,q)} \cdot p(post|blog)$$

results

blog distillation task of TREC Blog 2007

run	MAP	P@10	MRR
blogger	0.3272	0.4844	0.6892
posting	0.2325	0.3733	0.4850
TREC median	0.2035	-	-

conclusions

- expert retrieval models perform well on blog distillation.
 - the blogger model outperforms the posting model for the blog distillation task, for the expert finding task, the relative ranking is the other way around \Rightarrow there is a task difference:
 - for expert finding, a candidate is ranked highly when he is one of only few people mentioned in the context of the topic, while the amount of evidence or the number of other topics he is associated with are not (very) important.
 - for blog distillation, it appears we need to identify people that write mainly about the topic at hand.
- \Rightarrow we should explicitly model individual bloggers (as in the blogger model) and take a close look at the main themes that occupy them individually.

[1] K. Balog, L. Azzopardi, and M. de Rijke. Formal Models for Expert Finding in Enterprise Corpora. In Proc. SIGIR'06, pages 43-50, 2006. ACM Press.