Ranking Model Selection and Fusion for Effective Microblog Search

معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
Member of Qatar Foundation عضو مؤسسة قطر

Zhongyu Wei[1,2], Wei Gao[1], Tarek El-Ganainy[1], Walid Magdy[1], Kam-Fai Wong[2]

[1]Qatar Computing Research Institute, Doha, Qatar
[2]The Chinese University of Hong Kong, Hong Kong, China

1. Introduction

- **Our research studies how to improve re-ranking microblog search results by leveraging ranked lists from multiple rankers that are selected automatically**
- Our proposed solution:
  - **Query sensitive ranker selection**: Select *L* best performed rankers from candidate ranking models in a query-sensitive manner
  - **Result fusion of selected rankers**: Aggregate the ranked lists of selected rankers via different fusion techniques

2. Framework overview

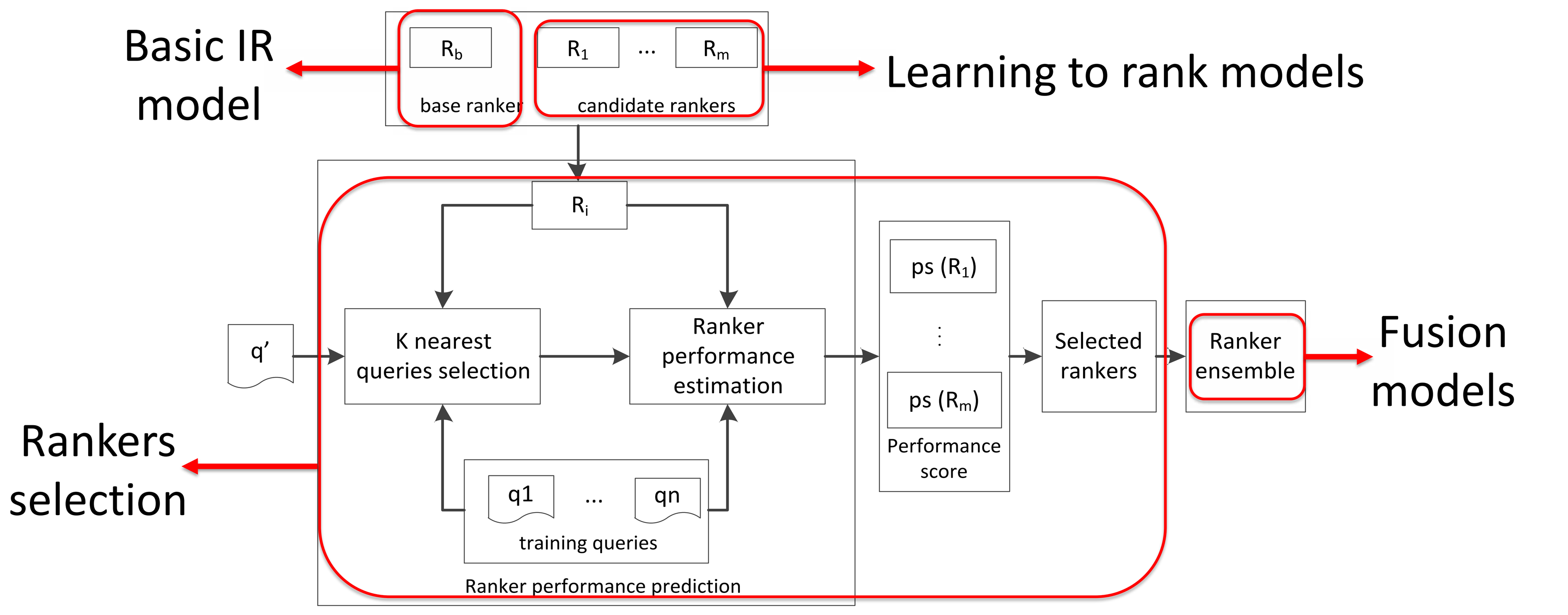
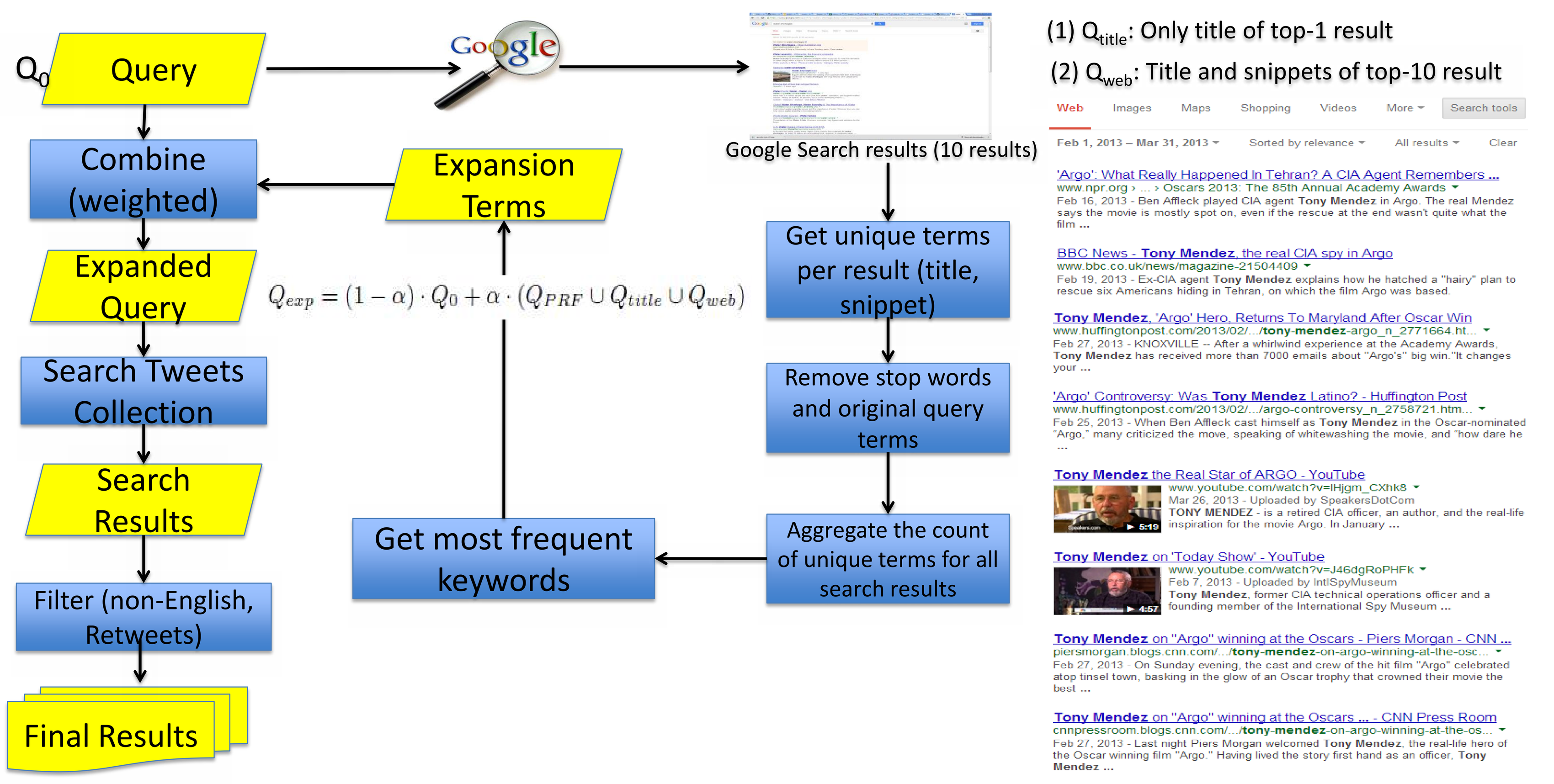


Figure 1. Our re-ranking framework

3. Base ranker: Web-search-based query expansion



4. Ranker selection

- Performance estimation for candidate ranker R_i given test query q':

$$\frac{1}{L}\sum_{l=1}^{L} ps(q^{(l)}, R_i)$$

Performance score of R_i on $q^{(l)}$: NDCG, MAP, MRR, etc..
$\{q^{(1)}, q^{(2)}, \ldots, q^{(L)}\}$ are L nearest neighbor training queries of q'

- L-Nearest neighbor queries of q':

$$sim(q', q^{(l)}) = -KLDiv(\mathbf{D}(R_i, q'), \mathbf{D}(R_i, q^{(l)}))$$

- Distribution of divergence between ranked lists of R_i and other rankers:

$$\mathbf{D}(R_i, q) = [D(R_b \| R_i, q), D(R_1 \| R_i, q), \ldots, D(R_m \| R_i, q)]$$

- Normalized divergence score between ranked lists of *two rankers:*

$$D(R_j \| R_i, q) = \frac{1}{Z}\sum_t |s_j(t) - s_i(t)|$$

5. Rank aggregation

- CombMNZ (Fox and Shaw, 1994):

$$CombMNZ(t) = |\{r \in R \mid rank_r(t) \le c\}| \times \sum_r score_r(t)$$

- Weighted Borda-fuse (Aslam and Montague, 2001)
- Weighted Condorcet-fuse (Montague and Aslam, 2002)
- Reciprocal rank fusion (RRF) (Cormack et al., 2009):

$$RRF(t) = \sum_{r \in R} \frac{1}{\kappa + rank_r(t)}$$

6. Candidate rankers

Algorithm	Validation %	Optimized metric
RankBoost (Freund et al., 2003)	--	--
MART (Friedman, 2001)	20%	--
RandomForest (Breiman, 2001)	20%	--
RankNet (Burges, et al., 2005)	20%	--
Coordinate Ascent (Metzler and Croft, 2007)	20%	P@30
		MAP
LambdaMART (Wu et al., 2010)	20%	P@30
		MAP

Table 1. Eight ranking models trained for ensemble

7. Datasets

Collection	# of tweets	# of terms	Average length
Tweets2011	16,141,809	155,562,660	9.64
Tweets2013	243,271,538	2,928,041,436	12.04

Table 2. The statistics of tweets collections

Query Set	# of queries	# of annotated	# of relevant
QS2011	50	40,855	2,864
QS2012	60	73,073	6,286
QS2013	60	71,279	9,011

Table 3. The statistics of relevance judgment

8. Experiments and Results

- **LM/LM_{webprf}:** baseline rankers
- **BestSingle:** Use best single ranker among all the candidate rankers
- **PMO:** Model selection method proposed by Peng et al. (2009) (top-1)
- **Best-sel:** Our model selection method (top-1)
- **x-all:** x-fuse of the results of all candidate rankers (without selection)
- **x-sel:** x-fuse of the results of selected candidate rankers (top-L)

	TREC2011		TREC2012		TREC2013	
	P@30	MAP	P@30	MAP	P@30	MAP
LM	0.4231	0.3897	0.3559	0.2329	0.4700	0.2731
BestSingle	***0.4673***	**0.4015**	***0.3887***	0.2399	0.4867	0.2803
PMO	***0.4633***	0.3873	*0.3814*	*0.2406*	0.4833	0.2641
CMNZ-all	***0.4510***	0.3621	*0.3814*	0.2326	0.4878	0.2664
Borda-all	***0.4483***	0.3646	*0.3819*	0.2362	0.4800	0.2708
Condorcet-all	***0.4633***	0.3925	***0.3870***	0.2391	0.4828	0.2759
RRF-all	***0.4558***	0.3672	***0.3915***	0.2376	0.4844	0.2732
Best-sel	***0.4633***	0.3940	***0.3859***	0.2337	*0.4900*	0.2746
CMNZ-sel	***0.4653***	0.3733	***0.3932******	0.2374	*0.4922*	0.2733
Borda-sel	***0.4633***	0.3940	***0.3859***	0.2337	0.4894	0.2669
Condorcet-sel	***0.4701***	0.3784	***0.3960******	*0.2406*	*0.4933*	0.2753
RRF-sel	***0.4633***	0.3940	***0.3927******	0.2363	*0.4917*	0.2622
LM_{webprf}	0.4905	0.4651	0.4356	0.2960	0.5350	0.3454
BestSingle	0.5075	0.4611	0.4514	0.2971	0.5494	*0.3559*
PMO	0.4966	0.4710	0.4452	0.2980	0.5567	0.3459
CMNZ-all	0.4884	0.4517	0.4452	0.2935	0.5589	0.3437
Borda-all	0.4932	0.4588	0.4345	0.2927	0.5600	0.3432
Condorcet-all	0.5048	0.4721	0.4463	0.2976	0.5494	0.3515
RRF-all	0.5014	0.4632	0.4492	0.2953	0.5561	0.3481
Best-sel	0.5102	0.4662	0.4531	0.3013	0.5561	0.3478
CMNZ-sel	***0.5143****	0.4656	*0.4571**	0.2976	*0.5700#*	0.3479
Borda-sel	*0.5102*	0.4670	0.4548	0.2977	*0.5678*	0.3493
Condorcet-sel	***0.5197*#$***	0.4671	***0.4605****	*0.3002$*	*0.5639*	0.3472
RRF-sel	*0.5136**	0.4575	0.4554	0.2999	*0.5695#*	0.3501
TREC Best System	0.4551	0.3350	0.4695	0.3469	0.5544	0.3506

Table 4. Re-ranking results based on LM and LM_{webprf}
(*Italic*: diff. with baseline p<0.05; **Bold**: diff. with baseline p<0.01; *: diff. with PMO p<0.05; #: diff. with BestSingle p<0.05; $: diff. between best x-sel and best x-all p<0.05; Underline: max value)

9. Conclusions

- Explore model selection and fusion methods for re-ranking microblog search results based on multiple learned ranking models
- An extension of query-sensitive model selection
- With a moderate base ranker, our re-ranking significantly outperforms the base ranker
- With a strong base ranker, our re-ranking significantly outperforms the best single ranker and the existing model selection method
- Simply aggregating all rankers is not better than the best single ranker