YJRS at the NTCIR-13 OpenLiveQ Task

Tomohiro Manabe Akiomi Nishida Sumio Fujita
Yahoo Japan Corporation
{tomanabe, anishida, sufujita}@yahoo-corp.jp

- We started from the baseline method.
- Our modifications are:
  - Extended BM25F as additional features
  - Five-fold cross validation
  - nDCG@10 as the objective function

Our Approaches

Baseline Method

- Linear combination of 77 features
  - 4 fields x 17 textual features, e.g. TF, LM, BM25, ⋯
  - 9 numeric features, e.g. # of answers/PVs, date, ⋯
- Weights are optimized by Coordinate Ascent (CA).

Extended BM25F as Ranking Features

- The BM25F is a non-linear function so that adding it as features may improve the model. (cf. neural net.)
  - Moreover we use numeric fields as well as TFs.

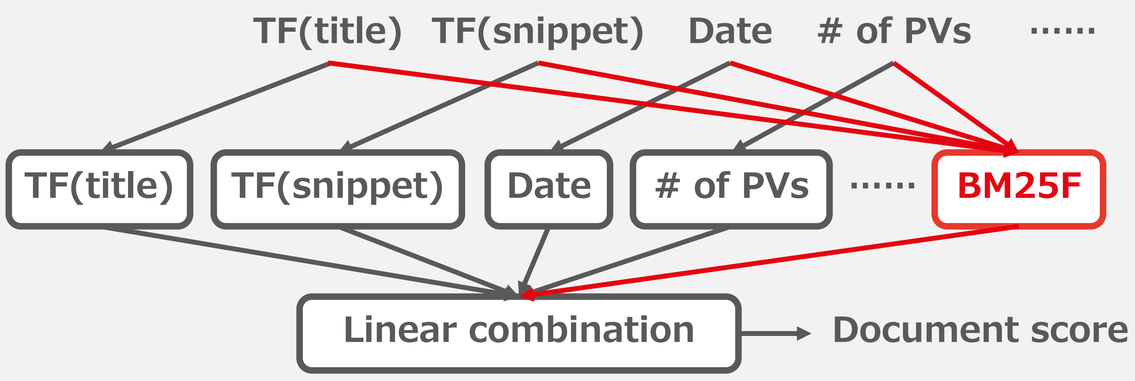
$$\sum_{t \in Q} \frac{w(t,D) + \alpha(D)}{k_1 + w(t,D) + \alpha(D)} \log \frac{N - \mathrm{df}(t) + 0.5}{\mathrm{df}(t) + 0.5}$$

k_1: Non-linear conversion

α: For adding numeric fields

$$\alpha(D) = \sum_{f \in D_N} v(f,D) \cdot boost_f$$

$$w(t,D) = \sum_{f \in D} \frac{\mathrm{tf}(t,f,D) \cdot boost_f}{(1 - b_f) + b_f \cdot \mathrm{len}(f,D)/\mathrm{avgLen}(f)}$$



- We tried 3 settings of BM25F. (Naïve: All fields, SERP: Fields on SERPs, SERP+: Fields prominent on SERPs)
- Adding the 3 settings as features more or less improved the offline score. (On nDCG@10, ~+10%)

Cross Validation

- Five-fold cross validation improved the offline score. (.380 → .412 on nDCG@10, +8.4%)

nDCG@10 as Objective Function

- Initially we used MAP as the objective function of CA.
  - Because quality of lower-ranked documents may be important in the greedy optimization process.
- Finally directly using nDCG@10 improved its score. (.396 → .419 on nDCG@10, +5.7%)

Evaluations

Feature Importance

- Setting 0.0 to the weight of each feature, we re-calculated nDCG@10 scores of the resulting rankings.
- The lower the score is, the more important the feature is.

Rank	nDCG	Feature
1	.193	Number of PVs
2	.2087	Log(Number of answers)
3	.2091	Number of answers
4	.210	Log(NormTF(Snippets))
5	.2111	Date of last modification

Rank	nDCG	Feature
6	.2112	Length of title
⋯	⋯	⋯
25	.2143	BM25F(SERP)
33	.2144	BM25F(Naive)
62	.215	BM25F(SERP+)

Offline Test Results

Rank	nDCG@10	Team
1	.445	OKSAT
2	.419	YJRS
3	.418	cdlab
4	.413	ORG
5	.406	Erler

Rank	ERR@10	Team
1	.276	OKSAT
2	.264	cdlab
3	.254	YJRS
4	.249	ORG
5	.245	Erler

Rank	Q-measure	Team
1	.713	YJRS
2	.707*	Erler
3	.702*	ORG
4	.700*	OKSAT
5	.697*	cdlab

- *: Statistically significant ($p < 0.05$) difference from YJRS based on Student's paired t-test
- Our run achieved the 2nd-best nDCG@10, 3rd-best ERR@10, and best Q-measure scores.
  - The differences on Q were statistically significant.

Online Test Results

Rank	Credit	Team
1	22.35k	Erler
2	22.31k	YJRS
3	21.3k*	ORG
4	20.0k*	cdlab
5	18.9k*	N-ANS

PVs we won	PVs we lost	Win-loss ratio	Team
35.9k	30.8k	.538*	Erler
40.5k	31.5k	.563*	cdlab
37.0k	28.5k	.565*	ORG
43.5k	24.7k	.637*	N-ANS
46.1k	24.8k	.650*	TUA1

- *: Statistically significant ($p < 0.05$) difference from YJRS based on the t-test/Pearson's chi-square test
- Our method achieved the 2nd-largest total credit.
  - Difference from the 1st was not stat. significant whereas one from the 3rd was.
- Our run consistently achieved the win-loss ratios better than 0.5 against all the other runs.
  - In stat. significantly larger number of PVs, our run won.

Conclusions

- Our method performed well due to its robustness.
- The BM25F is useful as learning-to-rank features.
- Well-known classical techniques, namely Coordinate Ascent and cross validation, are still useful.

NTCIR-13, December 05-08, 2017, NII, Tokyo, Japan