CS572IR HW2 Report

Xiangjue Dong (xdong57), Yanan Da (yda3)

March 2020

1 Implementation

For the Problem 1, we used the MART and LabmdaMART methods command -ranker 0 and -ranker 6 from RankLib on five folds separately and saved corresponding models by using -save commands as baseline models. Then we used -rank and -score commands to generate score files. We used the perl script with some modifications on parsing the prediction file to evaluate the baseline models on each fold. The results include NDCG@1-10, P@1-10, MeanNDCG, and MAP. Then we calculated the average cross the five folds.

For the Problem 2, according to the model feature statistics generated by RankLib, we can know only 41 of the 46 features were actually used. The minimum feature frequency was for feature 46, which was used only 7 times in the model. The maximum feature occurrence was for feature 45 which occurred 380 times in the model. To add some new features, we tried lots of different features, like the maximum, minimum, mean and sum of the term frequency, tf-idf, LMIR.JM, and so on. In addition, we put weights on some features, like length of URL. By choosing proper features, our new MART models trained on new data set with added features perform better than the MART baseline models in Problem 1 by the range of 0.31% to 9.34%. The NDCG and MAP comparison results with baseline models on test set are shown in Table 1. And other detailed results and files were generated as the same process as the Problem 1.

2 Results

Result files with NDCG and MAP results for each fold, and the average across folds, in the format like the LETOR baseline results are attached in the submission. The folder baseline_results contains baseline results using LambdaMart and MART from RankLib, and the folder newModel_results contains our models' results.

System	Performance	Improvement
MART-NDCG-MQ2007-Fold1 [baseline]	0.4652	
MART-NDCG-newMQ2007-Fold1	0.4822	+0.017 (+3.66%)
MART-NDCG-MQ2007-Fold2 [baseline]	0.4368	
MART-NDCG-newMQ2007-Fold2	0.4503	+0.0135 (+3.1%)
MART-NDCG-MQ2007-Fold3 [baseline]	0.4584	
MART-NDCG-newMQ2007-Fold3	0.465	+0.0066 (+1.44%)
MART-NDCG-MQ2007-Fold4 [baseline]	0.4109	
MART-NDCG-newMQ2007-Fold4	0.4493	+0.0384 (+9.34%)
MART-NDCG-MQ2007-Fold5 [baseline]	0.4431	
MART-NDCG-newMQ2007-Fold5	0.4562	+0.0131 (+2.97%)
MART-NDCG-MQ2008-Fold1 [baseline]	0.4931	
MART-NDCG-newMQ2008-Fold1	0.501	+0.0079 (+1.61%)
MART-NDCG-MQ2008-Fold2 [baseline]	0.4579	
MART-NDCG-newMQ2008-Fold2	0.4622	+0.0043 (+0.95%)
MART-NDCG-MQ2008-Fold3 [baseline]	0.4769	
MART-NDCG-newMQ2008-Fold3	0.4805	+0.0035 (+0.74%)
MART-NDCG-MQ2008-Fold4 [baseline]	0.5527	
MART-NDCG-newMQ2008-Fold4	0.5723	+0.0197 (+3.56%)
MART-NDCG-MQ2008-Fold5 [baseline]	0.5375	
MART-NDCG-newMQ2008-Fold5	0.5632	+0.0257 (+4.79%)
MART-MAP-MQ2007-Fold1 [baseline]	0.4767	
MART-MAP-newMQ2007-Fold1	0.4856	+0.0088 (+1.85%)
MART-MAP-MQ2007-Fold2 [baseline]	0.4607	
MART-MAP-newMQ2007-Fold2	0.4706	+0.01 (+2.16%)
MART-MAP-MQ2007-Fold3 [baseline]	0.4642	
MART-MAP-newMQ2007-Fold3	0.469	+0.0048 (+1.03%)
MART-MAP-MQ2007-Fold4 [baseline]	0.4342	
MART-MAP-newMQ2007-Fold4	0.4683	+0.0341 (+7.85%)
MART-MAP-MQ2007-Fold5 [baseline]	0.4613	
MART-MAP-newMQ2007-Fold5	0.477	+0.0157 (+3.41%)
MART-MAP-MQ2008-Fold1 [baseline]	0.4651	
MART-MAP-newMQ2008-Fold1	0.4799	+0.0147 (+3.17%)
MART-MAP-MQ2008-Fold2 [baseline]	0.4346	
MART-MAP-newMQ2008-Fold2	0.4359	+0.0013 (+0.31%)
MART-MAP-MQ2008-Fold3 [baseline]	0.4484	
MART-MAP-newMQ2008-Fold3	0.4549	+0.0064 (+1.43%)
MART-MAP-MQ2008-Fold4 [baseline]	0.5237	
MART-MAP-newMQ2008-Fold4	0.541	+0.0173 (+3.3%)
MART-MAP-MQ2008-Fold5 [baseline]	0.5019	
MART-MAP-newMQ2008-Fold5	0.5382	+0.0363 (+7.24%)

Table 1: Comparison results between new models and baseline models on testset