TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results
Group: (NEU) Northeastern University
Run ID: NEUEM1

Run type: Secondary
Description of run:
This is the EM algorithm for acquiring relevant judgements through crowdsourcing.

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	1	1986	43	26	0.054	0.979	0.021	0.946	0.383	0.650
416	1235	45	0	1183	7	45	0.011	0.994	0.006	0.989	0.432	0.623
417	2992	75	72	1303	1614	3	0.954	0.447	0.553	0.046	0.196	0.728
420	1136	37	27	371	728	10	0.724	0.338	0.662	0.276	0.464	0.524
427	1528	37	1	1484	7	36	0.039	0.995	0.005	0.961	0.260	0.605
432	2503	22	17	1002	1479	5	0.761	0.404	0.596	0.239	0.405	0.624
438	1798	162	1	1636	0	161	0.009	1.000	0.000	0.991	0.154	0.578
445	1404	60	0	1337	7	60	0.008	0.994	0.006	0.992	0.452	0.691
446	2020	156	29	1597	267	127	0.188	0.857	0.143	0.812	0.460	0.656
447	1588	16	16	575	997	0	0.971	0.366	0.634	0.029	0.186	0.652
Average	1826.000	63.700	16.400	1247.400	514.900	47.300	0.372	0.737	0.263	0.628	0.339	0.633

Table 1: This table shows per-topic statistics and overall averages for the run NEUEM1. The topics are 10 randomly selected topics from the TREC 8 ad-hoc task. A relevant document is positive and a non-relevant document is negative. The true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts are based on an adjudicated set of relevance judgments that differs from the original TREC-8 ad-hoc qrels. The true positive rate (TPR), false positive rate (FPR), true negative rate (TNR), and the false negative rate (FNR) are all smoothed values. Details of the computation of the logistic average misclassification (LAM) rate and the area under the curve (AUC) are given in the track overview paper. Some runs did not report a probability of relevance and thus will have NA for their AUC score.

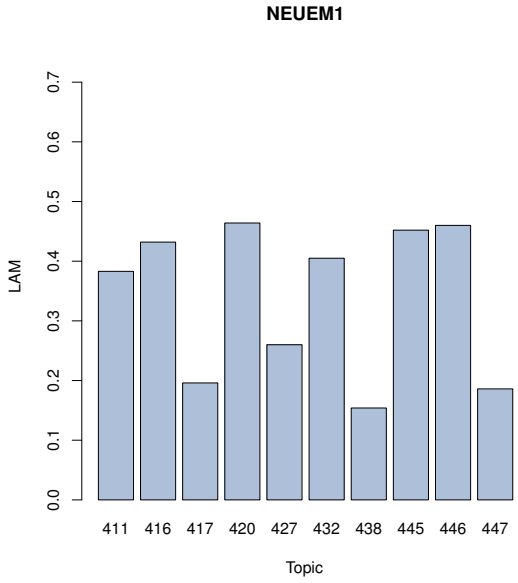

Figure 1: NEUEM1 LAM

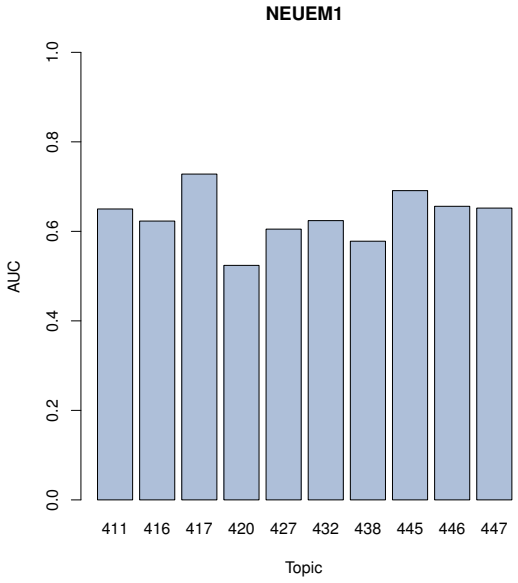

Figure 2: NEUEM1 AUC