TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results
Group: (NEU) Northeastern University
Run ID: NEUElo5

Run type: Secondary
Description of run:
We have used crowd sourcing to generate relevance judgements for pairs of documents in form of preference pair judgements. Elo rating algorithm is used to combine the preference judgements from crowds. The Elo rating system is a method for calculating the relative skill levels of players in two-player games such as chess.

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	4	2013	16	23	0.161	0.992	0.008	0.839	0.171	0.692
416	1235	45	9	1179	11	36	0.207	0.990	0.010	0.793	0.162	0.855
417	2992	75	8	2905	12	67	0.112	0.996	0.004	0.888	0.156	0.716
420	1136	37	9	1088	11	28	0.250	0.990	0.010	0.750	0.151	0.639
427	1528	37	10	1481	10	27	0.276	0.993	0.007	0.724	0.120	0.788
432	2503	22	0	2461	20	22	0.022	0.992	0.008	0.978	0.380	0.631
438	1798	162	31	1560	76	131	0.193	0.953	0.047	0.807	0.311	0.617
445	1404	60	7	1331	13	53	0.123	0.990	0.010	0.877	0.212	0.706
446	2020	156	14	1858	6	142	0.092	0.997	0.003	0.908	0.156	0.711
447	1588	16	8	1560	12	8	0.500	0.992	0.008	0.500	0.082	0.953
Average	1826.000	63.700	10.000	1743.600	18.700	53.700	0.194	0.988	0.012	0.806	0.190	0.731

Table 1: This table shows per-topic statistics and overall averages for the run NEUElo5. 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.

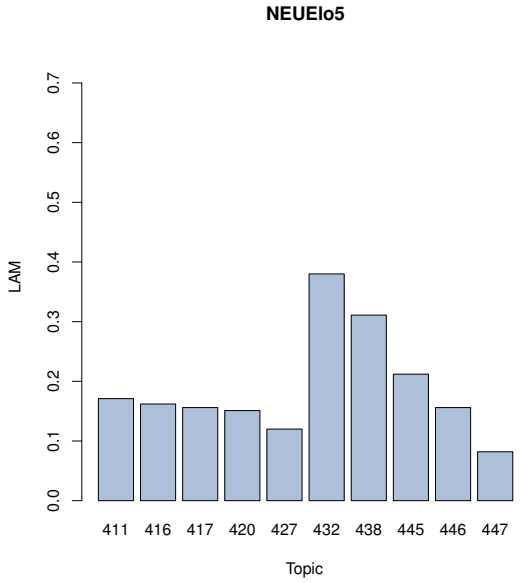

Figure 1: NEUElo5 LAM

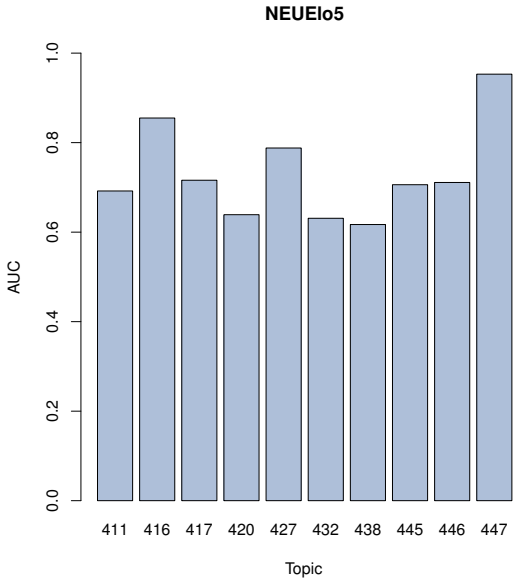

Figure 2: NEUElo5 AUC