TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results
Group: (NEU) Northeastern University
Run ID: NEUElo3

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	5	2022	7	22	0.196	0.996	0.004	0.804	0.110	0.690
416	1235	45	6	1177	13	39	0.141	0.989	0.011	0.859	0.209	0.745
417	2992	75	8	2896	21	67	0.112	0.993	0.007	0.888	0.195	0.695
420	1136	37	8	1092	7	29	0.224	0.993	0.007	0.776	0.134	0.598
427	1528	37	10	1478	13	27	0.276	0.991	0.009	0.724	0.134	0.756
432	2503	22	0	2475	6	22	0.022	0.997	0.003	0.978	0.256	0.598
438	1798	162	28	1516	120	134	0.175	0.926	0.074	0.825	0.380	0.583
445	1404	60	7	1341	3	53	0.123	0.997	0.003	0.877	0.120	0.783
446	2020	156	27	1831	33	129	0.175	0.982	0.018	0.825	0.227	0.720
447	1588	16	11	1572	0	5	0.676	1.000	0.000	0.324	0.012	0.953
Average	1826.000	63.700	11.000	1740.000	22.300	52.700	0.212	0.986	0.014	0.788	0.178	0.712

Table 1: This table shows per-topic statistics and overall averages for the run NEUElo3. 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.
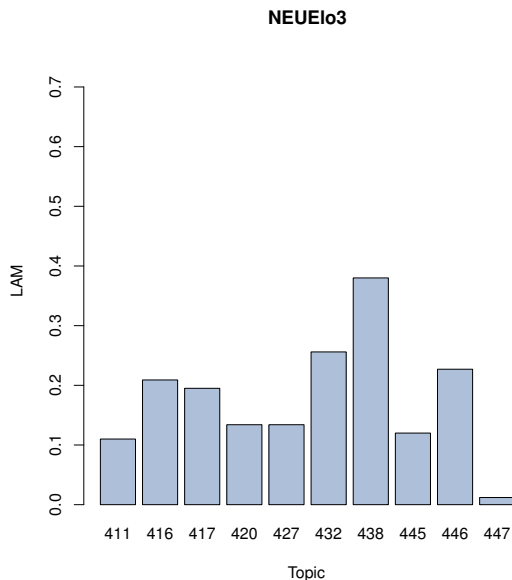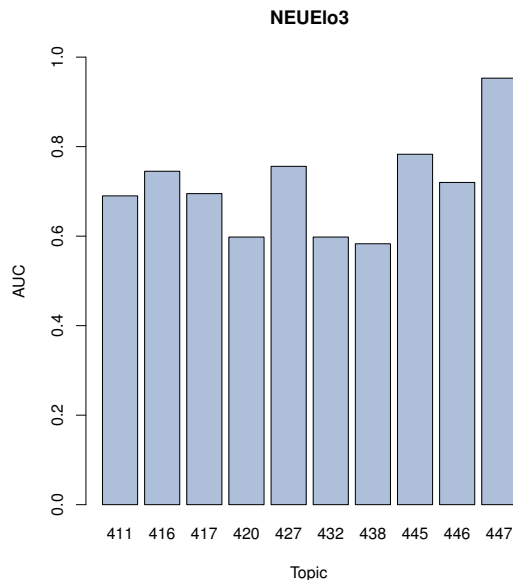

Figure 1: NEUElo3 LAM

Figure 2: NEUElo3 AUC