

# TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results

Group: (NEU) Northeastern University

Run ID: NEUElo4

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	6	2012	17	21	0.232	0.991	0.009	0.768	0.145	0.693
416	1235	45	7	1179	11	38	0.163	0.990	0.010	0.837	0.183	0.854
417	2992	75	9	2897	20	66	0.125	0.993	0.007	0.875	0.182	0.716
420	1136	37	10	1086	13	27	0.276	0.988	0.012	0.724	0.153	0.642
427	1528	37	11	1479	12	26	0.303	0.992	0.008	0.697	0.122	0.788
432	2503	22	0	2461	20	22	0.022	0.992	0.008	0.978	0.380	0.631
438	1798	162	39	1529	107	123	0.242	0.934	0.066	0.758	0.319	0.620
445	1404	60	6	1340	4	54	0.107	0.997	0.003	0.893	0.144	0.706
446	2020	156	25	1830	34	131	0.162	0.982	0.018	0.838	0.238	0.729
447	1588	16	12	1560	12	4	0.735	0.992	0.008	0.265	0.051	0.992
Average	1826.000	63.700	12.500	1737.300	25.000	51.200	0.237	0.985	0.015	0.763	0.192	0.737

Table 1: This table shows per-topic statistics and overall averages for the run NEUElo4. 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.

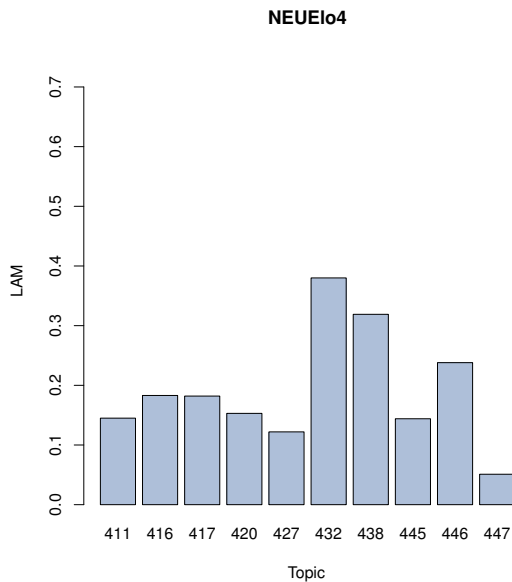


Figure 1: NEUElo4 LAM

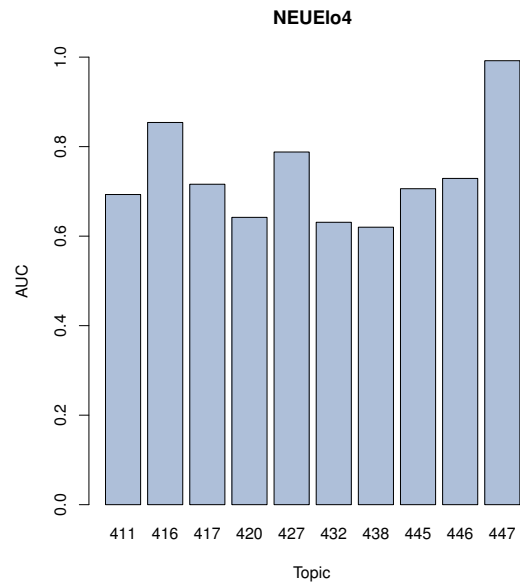


Figure 2: NEUElo4 AUC