

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

Group: (NEU) Northeastern University

Run ID: NEUElo2

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	1	2010	19	26	0.054	0.990	0.010	0.946	0.293	0.665
416	1235	45	4	1174	16	41	0.098	0.986	0.014	0.902	0.265	0.735
417	2992	75	3	2900	17	72	0.046	0.994	0.006	0.954	0.261	0.678
420	1136	37	3	1082	17	34	0.092	0.984	0.016	0.908	0.285	0.579
427	1528	37	8	1479	12	29	0.224	0.992	0.008	0.776	0.146	0.747
432	2503	22	0	2461	20	22	0.022	0.992	0.008	0.978	0.380	0.598
438	1798	162	8	1624	12	154	0.052	0.992	0.008	0.948	0.272	0.578
445	1404	60	6	1330	14	54	0.107	0.989	0.011	0.893	0.232	0.775
446	2020	156	3	1847	17	153	0.022	0.991	0.009	0.978	0.392	0.683
447	1588	16	1	1553	19	15	0.088	0.988	0.012	0.912	0.265	0.780
Average	1826.000	63.700	3.700	1746.000	16.300	60.000	0.080	0.990	0.010	0.920	0.279	0.682

Table 1: This table shows per-topic statistics and overall averages for the run NEUElo2. 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.

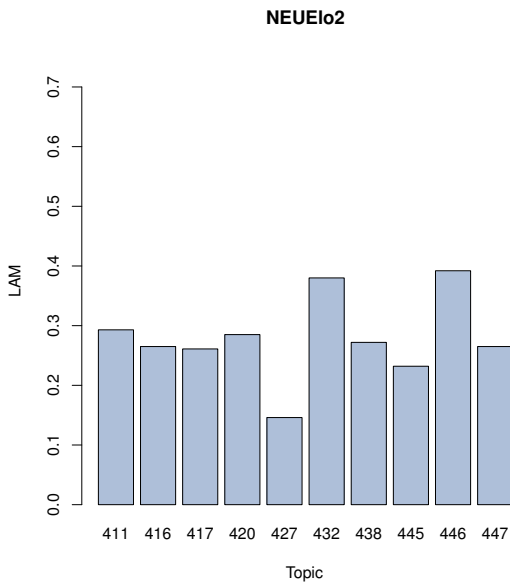


Figure 1: NEUElo2 LAM

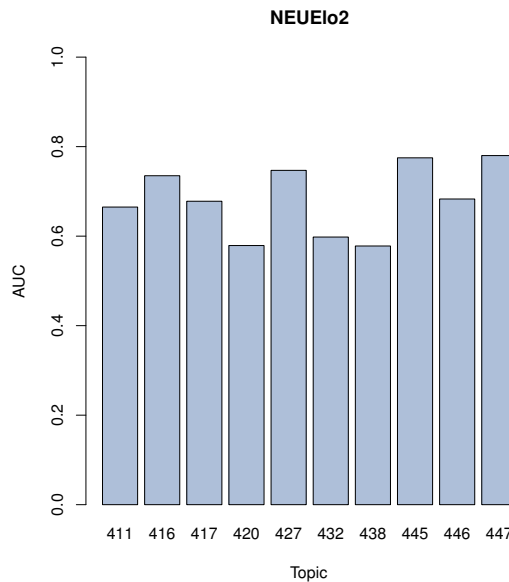


Figure 2: NEUElo2 AUC