

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

Group: (INFLBSTF) Stanford university

Run ID: INFLB2012

Run type: Primary

Description of run:

Our approach includes hybrid strategy using computer calculated results and crowdsourced results, and some interface issues on TRAT task.

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	0	2029	0	27	0.018	1.000	0.000	0.982	0.104	NA
416	1235	45	1	1190	0	44	0.033	1.000	0.000	0.967	0.100	NA
417	2992	75	4	2891	26	71	0.059	0.991	0.009	0.941	0.276	NA
420	1136	37	0	1099	0	37	0.013	1.000	0.000	0.987	0.156	NA
427	1528	37	1	1491	0	36	0.039	1.000	0.000	0.961	0.083	NA
432	2503	22	0	2481	0	22	0.022	1.000	0.000	0.978	0.087	NA
438	1798	162	14	1619	17	148	0.089	0.989	0.011	0.911	0.250	NA
445	1404	60	2	1344	0	58	0.041	1.000	0.000	0.959	0.085	NA
446	2020	156	2	1864	0	154	0.016	1.000	0.000	0.984	0.114	NA
447	1588	16	0	1572	0	16	0.029	1.000	0.000	0.971	0.093	NA
Average	1826.000	63.700	2.400	1758.000	4.300	61.300	0.036	0.998	0.002	0.964	0.135	NA

Table 1: This table shows per-topic statistics and overall averages for the run INFLB2012. 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.

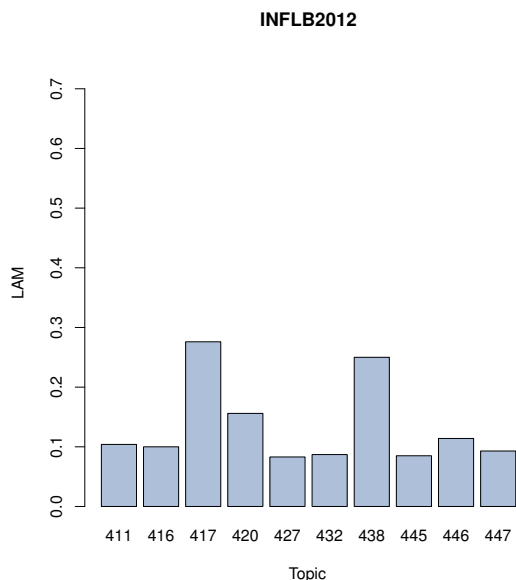


Figure 1: INFLB2012 LAM