

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

Group: (HAC) ECS, University of Southampton

Run ID: Orc2G

Run type: Secondary

Description of run:

Using topic analysis to select files to crowdsource, we obtained 2600 labels from Amazon Mechanical Turk workers. A simplified version of Independent Bayesian Classifier Combination was applied, learning from Topic features extracted from the text. Reliability of workers is also learnt from test examples and used to weight crowdsourced labels.

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	15	1787	242	12	0.554	0.881	0.119	0.446	0.249	0.807
416	1235	45	42	702	488	3	0.924	0.590	0.410	0.076	0.193	0.832
417	2992	75	58	2296	621	17	0.770	0.787	0.213	0.230	0.222	0.883
420	1136	37	25	696	403	12	0.671	0.633	0.367	0.329	0.348	0.635
427	1528	37	14	1153	338	23	0.382	0.773	0.227	0.618	0.408	0.616
432	2503	22	12	1797	684	10	0.543	0.724	0.276	0.457	0.361	0.691
438	1798	162	133	1069	567	29	0.819	0.653	0.347	0.181	0.255	0.810
445	1404	60	53	920	424	7	0.877	0.684	0.316	0.123	0.203	0.890
446	2020	156	73	1699	165	83	0.468	0.911	0.089	0.532	0.250	0.833
447	1588	16	8	1061	511	8	0.500	0.675	0.325	0.500	0.410	0.723
Average	1826.000	63.700	43.300	1318.000	444.300	20.400	0.651	0.731	0.269	0.349	0.290	0.772

Table 1: This table shows per-topic statistics and overall averages for the run Orc2G. 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.

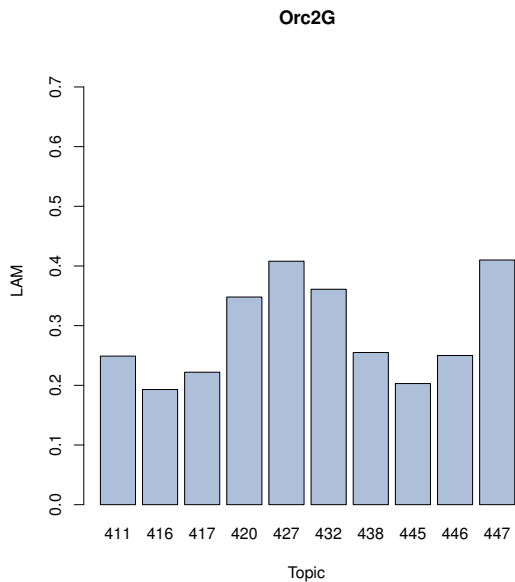


Figure 1: Orc2G LAM

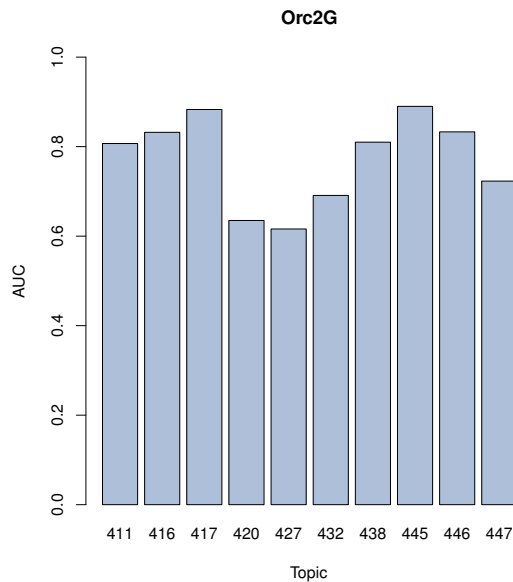


Figure 2: Orc2G AUC