TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results
Group: (HAC) ECS, University of Southampton
Run ID: OrcVBW80

Run type: Secondary
Description of run:
Using topic analysis to select files to crowdsource, we obtained 2600 labels from Amazon Mechanical Turk workers. Independent Bayesian Classifier Combination was applied, learning the reliability of individual crowd members and learning from Topic features extracted from the text. The crowd give estimates of their confidence when providing a label, which is taken into account by the classifier as it learns how to use the confidence label from test data.

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	22	1535	494	5	0.804	0.756	0.244	0.196	0.219	0.929
416	1235	45	40	840	350	5	0.880	0.706	0.294	0.120	0.192	0.859
417	2992	75	31	2483	434	44	0.414	0.851	0.149	0.586	0.332	0.734
420	1136	37	20	848	251	17	0.539	0.771	0.229	0.461	0.335	0.814
427	1528	37	18	1091	400	19	0.487	0.732	0.268	0.513	0.383	0.729
432	2503	22	15	1928	553	7	0.674	0.777	0.223	0.326	0.271	0.744
438	1798	162	113	1103	533	49	0.696	0.674	0.326	0.304	0.315	0.761
445	1404	60	36	893	451	24	0.598	0.664	0.336	0.402	0.368	0.743
446	2020	156	101	1466	398	55	0.646	0.786	0.214	0.354	0.278	0.831
447	1588	16	5	1208	364	11	0.324	0.768	0.232	0.676	0.443	0.814
Average	1826.000	63.700	40.100	1339.500	422.800	23.600	0.606	0.749	0.251	0.394	0.314	0.796

Table 1: This table shows per-topic statistics and overall averages for the run OrcVBW80. 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.
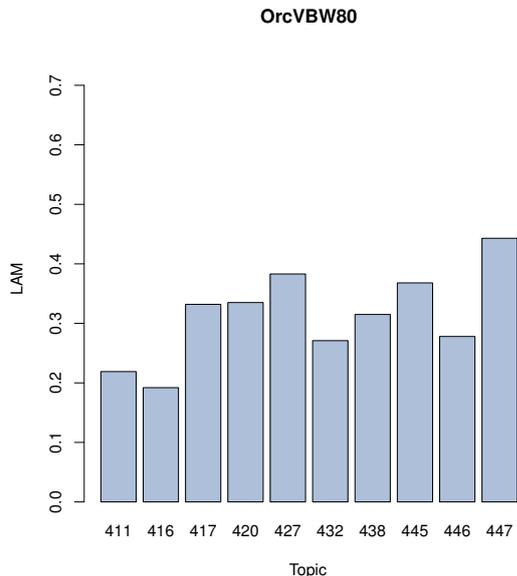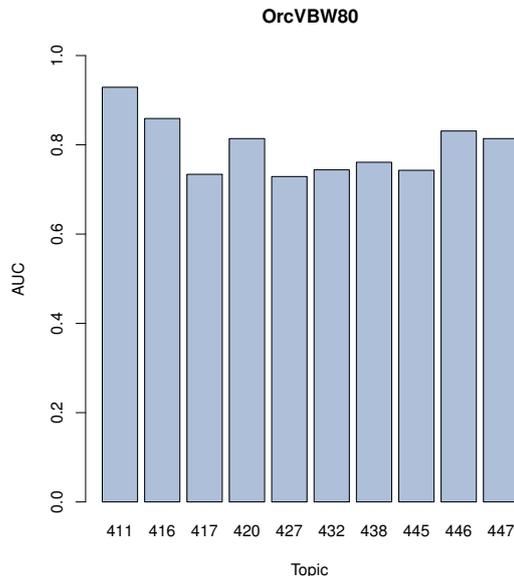

Figure 1: OrcVBW80 LAM

Figure 2: OrcVBW80 AUC