TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results
Group: (HAC) ECS, University of Southampton
Run ID: OrcVBW9Conf

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. This uses weaker priors so that worker reliability is inferred in a semi-supervised manner from the data.

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	20	1735	294	7	0.732	0.855	0.145	0.268	0.199	0.858
416	1235	45	44	689	501	1	0.967	0.579	0.421	0.033	0.135	0.853
417	2992	75	57	2146	771	18	0.757	0.736	0.264	0.243	0.254	0.813
420	1136	37	27	639	460	10	0.724	0.581	0.419	0.276	0.344	0.656
427	1528	37	5	1304	187	32	0.145	0.874	0.126	0.855	0.480	0.544
432	2503	22	17	1768	713	5	0.761	0.713	0.287	0.239	0.263	0.738
438	1798	162	118	972	664	44	0.727	0.594	0.406	0.273	0.336	0.774
445	1404	60	46	788	556	14	0.762	0.586	0.414	0.238	0.319	0.823
446	2020	156	132	1280	584	24	0.844	0.687	0.313	0.156	0.225	0.854
447	1588	16	5	1198	374	11	0.324	0.762	0.238	0.676	0.447	0.687
Average	1826.000	63.700	47.100	1251.900	510.400	16.600	0.674	0.697	0.303	0.326	0.300	0.760

Table 1: This table shows per-topic statistics and overall averages for the run OrcVBW9Conf. 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.
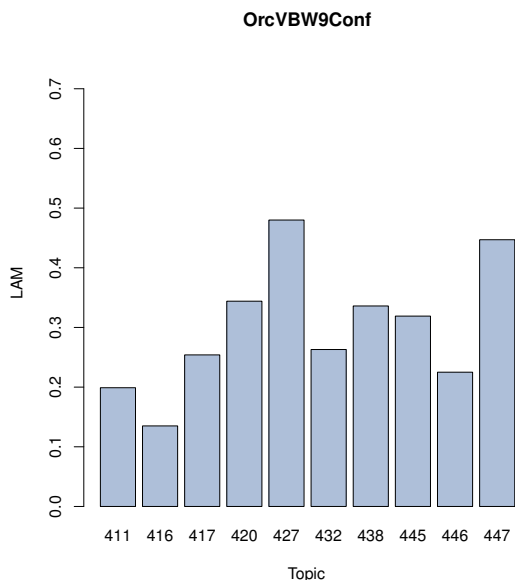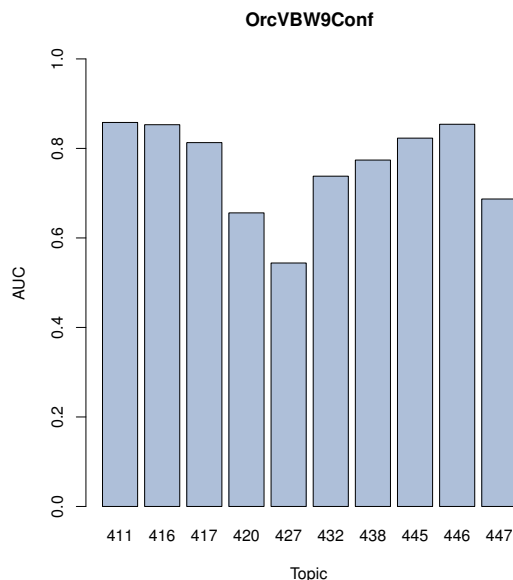

Figure 1: OrcVBW9Conf LAM

Figure 2: OrcVBW9Conf AUC