

TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results**Group: (HAC) ECS, University of Southampton****Run ID: Orc2GULConf**

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. Crowdsourced labels are taken into account when creating the final classification, in contrast to Orc2G, which only uses them to train a classifier. Confidence labels for the individual responses from the crowd are used to weight more confident responses more strongly when using the labels in classification.

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	18	1777	252	9	0.661	0.876	0.124	0.339	0.213	0.834
416	1235	45	42	724	466	3	0.924	0.608	0.392	0.076	0.187	0.846
417	2992	75	53	2328	589	22	0.704	0.798	0.202	0.296	0.246	0.725
420	1136	37	17	698	401	20	0.461	0.635	0.365	0.539	0.451	0.556
427	1528	37	14	1169	322	23	0.382	0.784	0.216	0.618	0.401	0.631
432	2503	22	13	1761	720	9	0.587	0.710	0.290	0.413	0.349	0.708
438	1798	162	130	1077	559	32	0.801	0.658	0.342	0.199	0.264	0.779
445	1404	60	44	915	429	16	0.730	0.681	0.319	0.270	0.294	0.742
446	2020	156	117	1362	502	39	0.748	0.731	0.269	0.252	0.260	0.812
447	1588	16	8	1073	499	8	0.500	0.682	0.318	0.500	0.406	0.720
Average	1826.000	63.700	45.600	1288.400	473.900	18.100	0.650	0.716	0.284	0.350	0.307	0.735

Table 1: This table shows per-topic statistics and overall averages for the run Orc2GULConf. 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.

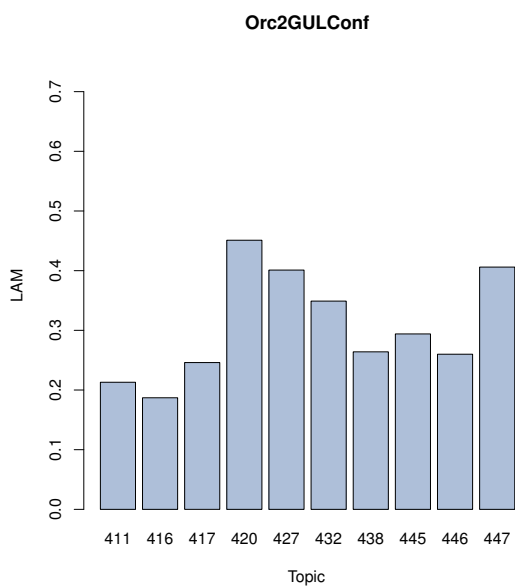


Figure 1: Orc2GULConf LAM

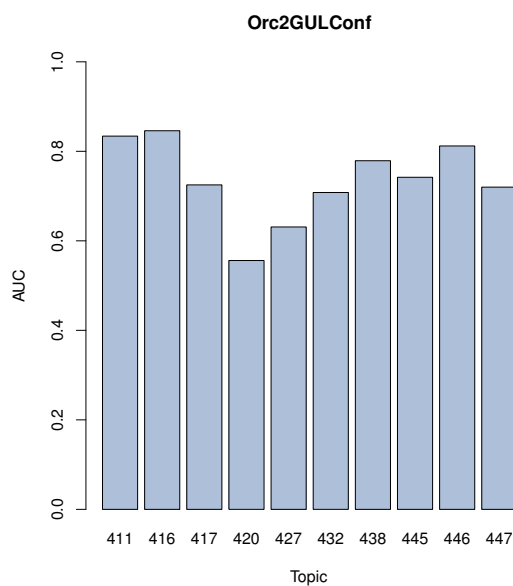


Figure 2: Orc2GULConf AUC