

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

Group: (HAC) ECS, University of Southampton

Run ID: Orc2GUL

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.

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	18	1782	247	9	0.661	0.878	0.122	0.339	0.211	0.832
416	1235	45	42	717	473	3	0.924	0.602	0.398	0.076	0.189	0.845
417	2992	75	49	2320	597	26	0.651	0.795	0.205	0.349	0.271	0.721
420	1136	37	21	701	398	16	0.566	0.638	0.362	0.434	0.398	0.566
427	1528	37	14	1167	324	23	0.382	0.783	0.217	0.618	0.402	0.631
432	2503	22	13	1755	726	9	0.587	0.707	0.293	0.413	0.351	0.710
438	1798	162	130	1073	563	32	0.801	0.656	0.344	0.199	0.266	0.780
445	1404	60	43	920	424	17	0.713	0.684	0.316	0.287	0.301	0.741
446	2020	156	117	1360	504	39	0.748	0.729	0.271	0.252	0.261	0.811
447	1588	16	8	1066	506	8	0.500	0.678	0.322	0.500	0.408	0.730
Average	1826.000	63.700	45.500	1286.100	476.200	18.200	0.653	0.715	0.285	0.347	0.306	0.737

Table 1: This table shows per-topic statistics and overall averages for the run Orc2GUL. 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.

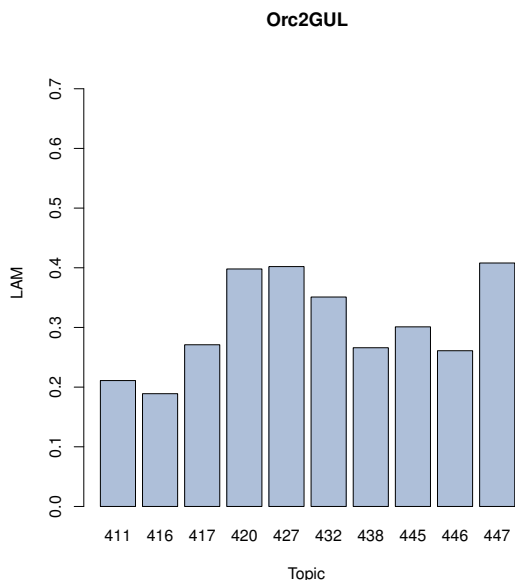


Figure 1: Orc2GUL LAM

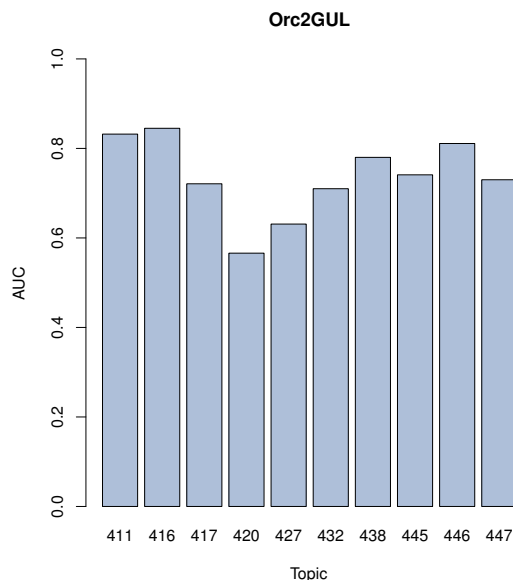


Figure 2: Orc2GUL AUC