

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

Group: (HAC) ECS, University of Southampton

Run ID: OrcVB1Conf

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, treating all crowd members as equal 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.

Results

| Topic | #Docs | #Rel | TP | TN | FP | FN | TPR | TNR | FPR | FNR | LAM | AUC |
|---------|----------|--------|--------|----------|---------|--------|-------|-------|-------|-------|-------|-------|
| 411 | 2056 | 27 | 21 | 1603 | 426 | 6 | 0.768 | 0.790 | 0.210 | 0.232 | 0.221 | 0.922 |
| 416 | 1235 | 45 | 43 | 738 | 452 | 2 | 0.946 | 0.620 | 0.380 | 0.054 | 0.158 | 0.858 |
| 417 | 2992 | 75 | 45 | 2231 | 686 | 30 | 0.599 | 0.765 | 0.235 | 0.401 | 0.312 | 0.866 |
| 420 | 1136 | 37 | 22 | 719 | 380 | 15 | 0.592 | 0.654 | 0.346 | 0.408 | 0.376 | 0.807 |
| 427 | 1528 | 37 | 10 | 1205 | 286 | 27 | 0.276 | 0.808 | 0.192 | 0.724 | 0.441 | 0.647 |
| 432 | 2503 | 22 | 15 | 1834 | 647 | 7 | 0.674 | 0.739 | 0.261 | 0.326 | 0.292 | 0.753 |
| 438 | 1798 | 162 | 126 | 956 | 680 | 36 | 0.776 | 0.584 | 0.416 | 0.224 | 0.312 | 0.798 |
| 445 | 1404 | 60 | 43 | 838 | 506 | 17 | 0.713 | 0.623 | 0.377 | 0.287 | 0.330 | 0.840 |
| 446 | 2020 | 156 | 134 | 1366 | 498 | 22 | 0.857 | 0.733 | 0.267 | 0.143 | 0.198 | 0.868 |
| 447 | 1588 | 16 | 5 | 1170 | 402 | 11 | 0.324 | 0.744 | 0.256 | 0.676 | 0.459 | 0.700 |
| Average | 1826.000 | 63.700 | 46.400 | 1266.000 | 496.300 | 17.300 | 0.652 | 0.706 | 0.294 | 0.348 | 0.310 | 0.806 |

Table 1: This table shows per-topic statistics and overall averages for the run OrcVB1Conf. 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.

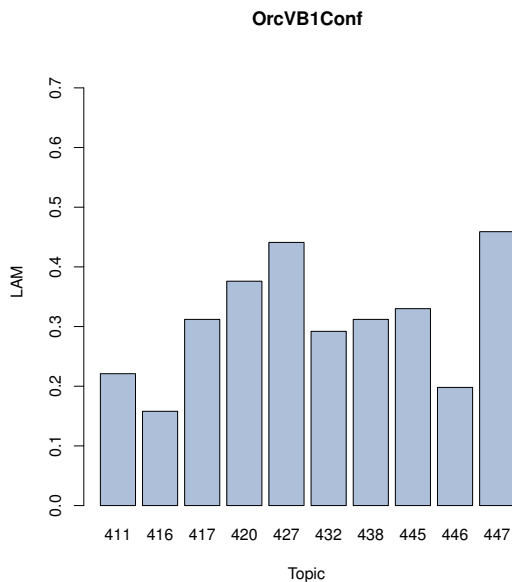


Figure 1: OrcVB1Conf LAM

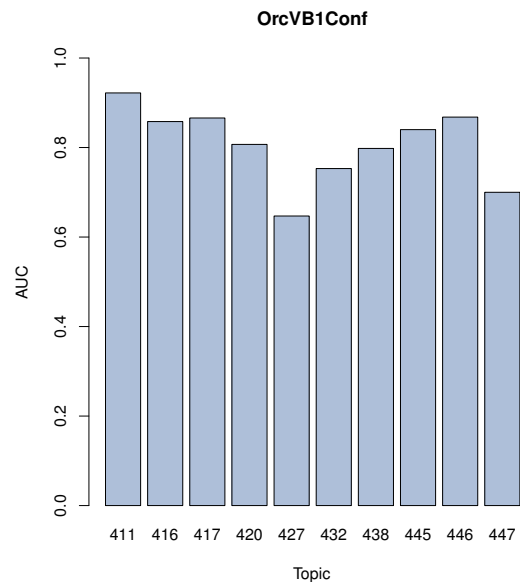


Figure 2: OrcVB1Conf AUC