

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

Group: (SetuServ) SetuServ

Run ID: SSECML2to99

Run type: Secondary

Description of run:

Post Error Correction (likely errors are identified using M/L for document with confidence scores between 2% and 99%), counting “not sures” as relevant documents

Results

| Topic | #Docs | #Rel | TP | TN | FP | FN | TPR | TNR | FPR | FNR | LAM | AUC |
|---------|----------|--------|--------|----------|--------|--------|-------|-------|-------|-------|-------|-------|
| 411 | 2056 | 27 | 21 | 1973 | 56 | 6 | 0.768 | 0.972 | 0.028 | 0.232 | 0.085 | 0.949 |
| 416 | 1235 | 45 | 41 | 1178 | 12 | 4 | 0.902 | 0.990 | 0.010 | 0.098 | 0.033 | 0.953 |
| 417 | 2992 | 75 | 60 | 2809 | 108 | 15 | 0.796 | 0.963 | 0.037 | 0.204 | 0.090 | 0.890 |
| 420 | 1136 | 37 | 34 | 1050 | 49 | 3 | 0.908 | 0.955 | 0.045 | 0.092 | 0.065 | 0.960 |
| 427 | 1528 | 37 | 31 | 1477 | 14 | 6 | 0.829 | 0.990 | 0.010 | 0.171 | 0.043 | 0.956 |
| 432 | 2503 | 22 | 9 | 2410 | 71 | 13 | 0.413 | 0.971 | 0.029 | 0.587 | 0.170 | 0.713 |
| 438 | 1798 | 162 | 151 | 1501 | 135 | 11 | 0.929 | 0.917 | 0.083 | 0.071 | 0.076 | 0.946 |
| 445 | 1404 | 60 | 49 | 1325 | 19 | 11 | 0.811 | 0.986 | 0.014 | 0.189 | 0.055 | 0.927 |
| 446 | 2020 | 156 | 97 | 1803 | 61 | 59 | 0.621 | 0.967 | 0.033 | 0.379 | 0.126 | 0.827 |
| 447 | 1588 | 16 | 15 | 1571 | 1 | 1 | 0.912 | 0.999 | 0.001 | 0.088 | 0.010 | 0.969 |
| Average | 1826.000 | 63.700 | 50.800 | 1709.700 | 52.600 | 12.900 | 0.789 | 0.971 | 0.029 | 0.211 | 0.075 | 0.909 |

Table 1: This table shows per-topic statistics and overall averages for the run SSECML2to99. 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.

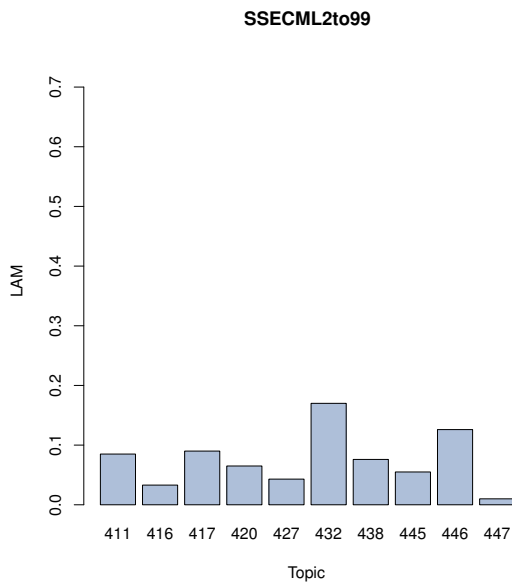


Figure 1: SSECML2to99 LAM

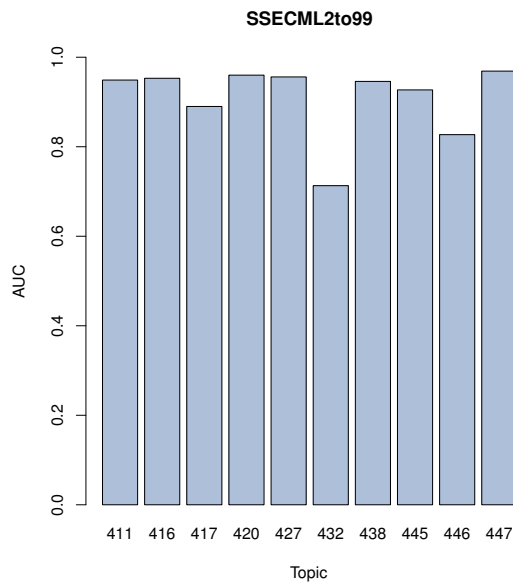


Figure 2: SSECML2to99 AUC