



A Tree in a Random Forest

Junzhe Zhu

We web-scraped 450000+ comments from New York Times, applied a LDA model to analyze their topic distribution, and manually marked these topics. We then derived a Random Decision Tree Forest of 100 regression trees, with more than 1100 nodes in each tree, in order to infer audience engagement(indicated by comment length) from topic distribution for each comment. The visualization of the tree helped us determine the importance of each topic in terms of how it affects audience's attention.