Learning to Classify Email into Speech Acts

It is often useful to classify email according to the intent of the sender (e.g., "propose a meeting", "deliver information"). We present experimental results in learning to classify email in this fashion, where each class corresponds to a verb-noun pair taken from a predefined ontology describing typical “email speech acts”. We demonstrate that, although this categorization problem is quite different from “topical” text classification, certain categories of messages can nonetheless be detected with high precision (above 80%) and reasonable recall (above 50%) using existing text-classification learning methods. This result suggests that useful task-tracking tools could be constructed based on automatic classification into this taxonomy. We also show that shallow linguistic analysis of emails can improve classification performance, and we present an analysis of the performance gains likely to result from sequential classification methods.