

## Abstract

**T**HIS work describes an application of Latent Dirichlet Allocation (LDA) text modeling to find clusters in various text corpora. Specific optimizations and applications are made to model medical data consisting of patient observations. The LDA topic model is first explored with a set of labeled (supervised) data to evaluate performance and demonstrate the viability of an unsupervised system. The supervised data is clustered and this reduced representation is used to classify the documents. The class labels are then compared to the original labels to evaluate the topic model. The performance on the unsupervised set is evaluated through observation of the clusters and the most-likely topic distributions, as well as the log-likelihood score of the topic model.