

Abstract

This thesis explores a new approach for classifying documents' polarity as well as discovering the hidden semantic structure in documents. While most existing methods in sentiment analysis require sentiment labels of training documents, the approach proposed in this thesis is fully unsupervised by leveraging probabilistic topic modeling techniques. The model proposed in this thesis is built upon joint sentiment topic model (JST), which is derived from the foundation model in topic modeling, latent Dirichlet allocation (LDA). Because the results of JST are sensitive to the initial priors parameters, the hierarchical priors structure is applied to improve the performance of JST. This new hybrid model is tested on a dataset consisting of product reviews from Amazon, and it yields higher classification accuracy and better quality topics than the original model.