Title

Textual Feature Extraction for Sentiment Analysis

Introduction

In this project, we perform the sentiment analysis over the IMDB movie reviews. We classified movie reviews to either of the general attitudes: positive and negative. For this supervised learning process, the crucial part is the feature extraction from the text reviews. We compared two state-of-art models of text analysis using two popular classifiers. One is the one-hot key binary word vector which is widely used to encoding text in traditional natural language processing (NLP), while the other is the Doc2Vec technique where document vectors are learned via artificial neural network. These models are finally evaluated in metrics of confusion matrix and accuracy of binary classifications.

My Contribution

* Implement the feature extraction models using Python package gensim.
* Implement the classification and evaluation procedures using Scikit-learn APIs.
* Data visualization using python Matplotlib package.

Reference