Practice: Sentiment analysis

This dataset was created for the paper ‘From Group to Individual Labels using Deep Features’, Kotzias et. al,. KDD 2015. It contains sentences labeled with positive or negative sentiment, extracted from reviews of products, movies, and restaurants. The sentences come from three different websites/fields: imdb.com, amazon.com, yelp.com.

For each website, there exist 500 positive and 500 negative sentences. Those were selected randomly for larger datasets of reviews. Score is either 1 (for positive) or 0 (for negative).

For the purpose of this practice we have compiled the data in the file `sa.csv`.

We propose the following questions:

1- Read the data.

2- Tokenize the sentences being max sentence length 25 and 5000 max number of words to include.

3- Separate the data into 75% training and 25% test.

4- Using a trainable embedding with dimension equal to 20. Fit a model with a hidden dense layer (10 units) to predict - sentences sentiment.

5- Using a trainable embedding with dimension equal to 20. Fit a model with a RNN hidden layer (28 units) to predict sentences sentiment.

6- Applying a pre-trained embedding with the 50-dimensional embedding vectors from GloVe, fit the model defined in 5) to predict sentences sentiment.