

CS4361/5361 Machine Learning

Fall 2020

Text Classification

The program `read_sentences.py` reads sentences from online classic books and converts them to a list of sentences, where each sentence is a list of words.

1. Write a function that receives a sentence and returns a 2D array containing the embeddings of the words in the sentence. Your function should receive the embeddings dictionary, the sentence and the desired length of the representation; if the sentence is shorter than the desired length, pad the array with zeros; if it's longer, truncate the representation.
2. Apply the function to produce an embedding representation of each of the sentences in the three books used in the `read_sentences.py` program and generate a dataset containing examples of 3 classes, one for each book.
3. Randomly split the data into training and testing.
4. Train and test a system to determine the book each sentence belongs to.