



IMapBook collaborative discussions classification

Marjan Bogataj, David Peklenik Urbanč, Aljaž Dobnik

Abstract

Advisors: Slavko Žitnik

Introduction

The idea behind the project is to classify messages into their respective categories. The messages were extracted from an app for book club discussions for elementary schools. The pupils read one of three books, and were asked questions regarding the contents of the books. Each group had an internal discussion which resulted in a group response to the question asked. In this paper we will analyze their internal discussion and implement a natural language processing algorithm for message classification.

Methods

Firstly, we will split the data based on a specific topic and then based on book clubs to get separate group discussions.

Then we will implement text pre-processing, which will include stop word deletion, text to lowercase, removal of all digits and extra spaces (not special characters, since they can represent emotions with emoticons). We will test different text representations (Word2vec, TF-IDF) and different models (Multinomial Naive Bayes, SVM, Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)). To evaluate our models, we will try different evaluation techniques (Precision/Recall, F score).