11-747 Assignment 1: Text Classifier

Huiming Jin
huimingj@andrew.cmu.edu

Abstract

In this assignment, we implement multi-class text classifiers from scratch. Besides the most basic CNN model, various deep neural architectures are tried, as well as several machine learning strategies for better performance. With acceptable time and memory spending, the best model achieves 86.6% accuracy on the given validation set. We also conduct a series of experiments to analyze our models.

1 Introduction

For this assignment, we are given simply preprocessed multi-class text classification task data. Each instance in the data is a sentence along with a single label that indicates the topic of sentence.

Label	Sentence
Music	The song was written by Madonna and produced by Benitez .
Video games	The developers of the game added round metal shoulders on Samus ́varia Suit to differentiate it from her Power Suit , since both looked similar on the Game Boy ś limited greyscale display .
Sports and recreation	The series went to seven games .

Table 1: Some data examples.

In the submission of this assignment, the predicted labels for the validation set and the test set are listed in $\texttt{predict}_v al.txtand$