Neural Networks for Sentiment Analysis

Project Overview

In the project, FFNN and RNN have been implemented and compared for 5-class sentiment analysis. The models are trained and tested on a Yelp review dataset to predict the sentiment score (1-5) for each review.

Dataset

Training Set: This includes labeled reviews for training the models.

Validation Set: This set is used for tuning hyperparameters and doing intermediate evaluations.

Test Set: This set will be used to perform the final evaluation of the performance of the model.

Word Embeddings: It's a pretrained embedding that shall be used in order to initialize the RNN model.

Models

FFNN

Input reviews - these will be provided as fixed vectors.

Carry out a hidden layer and output layer, and apply softmax for classification.

RNN

Input reviews - these will be provided as sequences of word embeddings.

Carry out recurrent layers for the representation of sequences and classify using softmax.

Files and Directories

- `ffnn.py`: FFNN model implementation

- `rnn.py`: RNN model implementation

- `training.json`, `validation.json`, `test.json`: input files used for training, validation and testing

- `word_embedding.pkl`: pre-trained word embeddings

- `CS4375_HW_2_writeupEnesOzturk.pdf`: project report that includes details of implementation and models' performance.

- `FFNN_and_RNN_Output_Analysis.txt`: models' performances analysis

Directions

1. Install the necessary python dependencies from `requirements.txt`.

2. Run either `ffnn.py` or `rnn.py` with appropriate arguments to train the models.

   Example:

python ffnn.py --hidden_dim 64 --epochs 10 --train_data training.json --val_data validation.json

   ```

3. Run the models on the test set.

4. Go through the analysis and results presented in the report and output files.

Requirements

- Python 3.8

- PyTorch 1.10.1

- Other libraries - see `requirements.txt`.

Acknowledgements

This is a course project for CS4375 on neural networks.