PROJECT STATEMENT

Topic

Sentiment Analysis of Movie Reviews Using LSTM

Introduction

You are asked to modify a Deep Learning structure, consisting of a word embedding layer, a LSTM layer and a classification layer, to perform sentiment classification on movie review domain.

Environments

PyTorch: https://pytorch.org/

GloVe: https://nlp.stanford.edu/projects/glove/

- I strongly recommend, also recommended by *pytorch.org*, to use *Anaconda* as the package manager since it installs all dependencies.
- You may also need other prerequisite packages, such as *numpy*, and you can simply use *pip install numpy* to install them under *Anaconda*.
- The package on e-learning is tested using Windows (OS), Python 3.7 (Language), None (CUDA).
- The package uses *glove.840B.300d.txt* as the pre-trained GloVe word vectors

Package

The package contains the following elements:

- Dataset
  - GloVe
    - Glove.840B.300d.txt # Not included, you need to download it from internet, unzip it and add it to the directory.
  - Stsa
    - Label.dev # validation labels
    - Label.test # testing labels
    - Label.train # training labels
    - S1.dev # validation data
    - S1.test # testing data
    - S1.train # training data
- Savedir
  - Model.pickle # save the best model on validation data
- Data.py # data processing
- Models.py # main structure of the model (need modification)
- Multis.py # not important to the project ^_^
- Train_nli.py # training process of the model (need modification)

What you need to do for the project

- Install the environment and execute the package (easy)
- Fine-tune the hyperparameters and show their influence on the experimental results (easy)
  - Modify line 17 to 36 in *train_nli.py* to change the hyperparameters and show the experimental results after the modification.

- Make structure-level modification on the model (Medium)
    - Modify line 41 to 62 in *models.py* and corresponding hyperparameters in *train_nli.py*. Try to add/delete linear layers or add *dropout* to the linear layers and show the experimental results after the modification.
- Make modifications on the Deep Neural Network Model (Hard)
    - Modify line 7 to 35 in *models.py* and corresponding hyperparameters and structures in *train_nli.py* and *models.py*. Try to extend the LSTM to Bi-LSTM and show the experimental results after the modification

Sample Report

Have you successfully installed the environment and execute the package?

- Yes (60 points)
- No (50 points). If *No*, please analyze the reasons.

Have you made modifications on the hyperparameters?

- Yes (10 points). If yes, list the parameters you have changed, show the new experimental results and your analysis.
- No

Have you made structure level modifications of the model?

- Yes (10 points). If yes, list the modifications, show the new experimental results and your analysis.
- No

Have you extended the LSTM model to Bi-LSTM model?

- Yes (10 points). If yes, show the new experimental results and your analysis.
- No

What to submit

The updated package with your report. Please include the report in the package. The report worth 10 points of this project.

Something to pay attention to

Except from the report, please DO NOT add/delete files or change the names of the files in the package. All the modifications should be done within the package, specifically on *train_nli.py* and *models.py* files and no other files need to be modified.

DO NOT include the word vectors (*glove.840B.300d.txt*) in the package when you do the submission. The size of the word vectors is almost 6G, you cannot submit the package to e-learning if you include it. I will add the word vectors for you when testing your project.

Your report should be consistent with the package. DO NOT include any information that does not appear in the package into your report.

The package will be compiled and executed by the command line **python train_nli.py**

Contact TA Linrui Zhang (lxz132230@utdallas.edu) for further questions.