

Gilmore Girls Dialogue Generator Documentation

Author: Yang Yang

Date: 12/9/2018

- 1) An overview of the function of the code (i.e., what it does and what it can be used for).

The project contains four main parts:

- A. Scraped training data (all seasons TV scripts from web)
- B. Trained a character-level LSTM neural net model using python library Keras and tensorflow GPU backend on Gilmore Girls TV scripts by actors (Rory vs Lorelai)
- C. Saved the model object and weights to make prediction from user input using 40-character sequence
- D. Developed a Web-UI to display prediction results and allows interactions through button react component.

This Web-UI can be used to generate texts (length between 100 to 400 characters) given user input sequence. It can be used to compare results from varying diversity parameter, as well as compare results from the two actors using the same seeding sequence.

- 2) Documentation of how the software is implemented with sufficient detail so that others can have a basic understanding of your code for future extension or any further improvement.

I will describe in detail each of the four parts mentioned above here.

- A. The code for web scraping of the Gilmore Girls script is included in the \scrape folder. The output file is .jl and can be read in as a json file.
- B. The LSTM model training and prediction code is save in the \src folder.
 - a. I subset the dataset by the two actors (Rory and Lorelai) and created character level corpus using all the lines by each of them. The two corpuses are very similar but not 100% the same. For simplicity I separated the modelling part by the two actors into two code files (**lstmmmodel_lorelai.py** and **lstmmmodel_rory.py**)
 - b. The model training employed tensorflow GPU and saved the two model weights objects as **weights_lorelai.hdf5** and **weights_rory.hdf5** in the \src folder.
 - c. The **predict.py** code will take user input (40 characters) or generate one if you don't want to type in, converts this string into the required shape format for the LSTM model and output the next character based on a probabilistic distribution.
 - d. Navigate to the \ui folder from command line, execute **unicorn server:app** (from command line) and open any local browser and type **localhost:8000** to interact with the web UI.

- 3) Documentation of the usage of the software including either documentation of usages of APIs or detailed instructions on how to install and run a software, whichever is applicable.

Step 1: Create a new conda environment (recommended)

After you have cloned the project, I recommend you create a new environment before installing the required python libraries.

Go to your terminal and run **conda create -n myenv**

Note: replace **myenv** with the environment name you prefer

Activate the new environment:

Windows: **activate myenv**

macOS and Linux: **source activate myenv**

NOTE: Replace **myenv** with the name of the environment.

Step 2: Install the following using pip

- a. tensorflow
- b. keras
- c. falcon
- d. gunicorn
- e. numpy
- f. pandas

Note: the model is trained using GPU device but already saved the model object in the \src. You can run the prediction and the app using CPU tensorflow.

Step 3: Navigate to the ui subdirectory and execute in command line

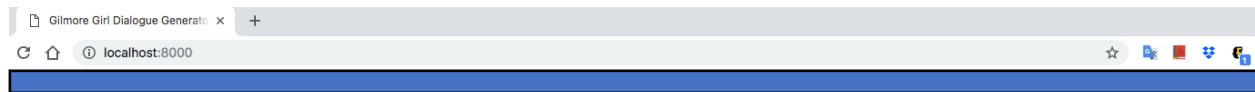
gunicorn server:app

You should see the following:

```
my's-MacBook-Pro:ui ayang$ gunicorn server:app
[2018-12-14 15:11:19 -0600] [82016] [INFO] Starting gunicorn 19.9.0
[2018-12-14 15:11:19 -0600] [82016] [INFO] Listening at: http://127.0.0.1:8000 (
82016)
[2018-12-14 15:11:19 -0600] [82016] [INFO] Using worker: sync
[2018-12-14 15:11:19 -0600] [82019] [INFO] Booting worker with pid: 82019
```

Step 4: Open any browser, and go to '**localhost:8000**' to access the UI

You should be able to see the UI show up as below



Follow these steps:

- Select your favorite Gilmore Girl from the **Role** dropdown list
- Input any 40-character long **leading sequence/partial sentence**
- Or simply hit the **Generate Seq** button to generate random sequence
- Input a number between 100 and 400 for the **output length** (default 200 character)
- Hit **Submit** and wait for your result

Note: the leading seed will be displayed before the generated text.

Are you ready?

Step 5: Follow the instructions on the web app to explore the results you generated.

Example output:

Role: **Rory** Leading Seq Input: **de, i assure you, i was** **Generate Seq** Output Length: **300** **Submit**

Follow these steps:

- Select your favorite Gilmore Girl from the **Role** dropdown list
- Input any 40-character long **leading sequence/partial sentence**
- Or simply hit the **Generate Seq** button to generate random sequence
- Input a number between 100 and 400 for the **output length** (default 200 character)
- Hit **Submit** and wait for your result

Note: the leading seed will be displayed before the generated text.

Success!

de, i assure you, i was devoid of the chance of the paper. i don't know what i want to do that. i don't know. you can go to the car. i don't know. i don't think i don't know what you don't know what you don't think you don't know what i was a first of the table and the way on the stuff here. i don't know what i was gonna go to the boo (Diversity: 0.2)

de, i assure you, i was devoid of the charation. you know what? what's that, but they're so not in the bed. nothing what to staying the but content and to stop you? he's an all the concered. grandma took my friends. that was may better of anything about you to all this of this? mom, that is good to the packed out of the whole time. th (Diversity: 0.5)

de, i assure you, i was devoid of the charoas on as i'm spaybning that the arcousedmalls issolou be a gone anthur and meper chinates that waits with my will come on this. okay. no, mom. woo, i have to call there. oh, maybe loys out. huh? you said what, it's thin we have the knewen the but buballv an a recottaed. okav? oh okav. ves. i fe (Diversity: 1.0)

There are a few parameters you can change and compare the outputs across:

- Role/Actor
- Input sequence/ generated sequence
- Output length

