**Task Name**: Generating lyrics like Ed-Sheeran songs

**Team Members**: Jayanth, Sai Priya

**Data Pre-processing**

* Removing empty songs
* Data de-duplication
* Converting songs to lowercase
* Removing special characters
* Removing stop words

**EDA**

* Total unique songs
* Median length of songs (characters, words)
* Top words using word cloud

**----MODEL BUILDING----**

**Technique used**

* For character-based model: Predict a character after a sequence of 50 characters
* For character-based model: Predict a word after a sequence of 50 words

**Splitting the dataset:** Training set (80%), Test set (20%)

**Networks used**: RNN, Bi-directional RNN, Multi-layer RNN

**Embeddings used**

* Character-wise one hot encoding
* Word-wise one hot encoding
* Word2Vec embedding (fine-tuned)

(Pre-trained embeddings based on Wikipedia (100 dimensions per word))

<https://wikipedia2vec.github.io/wikipedia2vec/pretrained/>

**Optimizer:** Adam

**Hyper-parameter combinations tested**

|  |  |  |
| --- | --- | --- |
| **Learning rate** | **Batch size** | **Epochs** |
| 0.01 | 128 | 5,10,15,20,25,30 |
| 0.01 | 256 | 5,10,15,20,25,30 |
| 0.001 | 128 | 5,10,15,20,25,30 |
| 0.001 | 128 | 5,10,15,20,25,30 |

**Note:** High batch size, less epochs and less combinations are chosen due to time constraints for model building

**Goodness of model**

BLUE score was calculated with reference as validation dataset

* BLEU score (1-gram)
* BLEU score (2-gram)
* BLEU score (3-gram)
* BLEU score (4-gram)

**Other Findings**

* Time for training model
* Plot for BLEU score (2-gram) Vs Epochs
* Plot for BLEU score (2-gram) Vs Batch size

**Best Parameters**

Best parameters were found using BLEU score, as the current task suggests using BLEU score (usually identified based on validation loss and validation accuracy)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Network** | **Embedding** | **Learning Rate** | **Batch Size** | **Epochs** | **BLEU score (2-grams)** | **Time (min)** |
| RNN | Word | 0.01 | 256 | 20 | 0.69 | 2.97 |
| RNN | Word2vec | 0.001 | 128 | 20 | 0.23 | 8.57 |
| Bi-RNN | Word | 0.01 | 128 | 30 | 0.56 | 15.63 |
| Bi-RNN | Word2vec | 0.001 | 128 | 30 | 0.38 | 22.91 |
| Multi-RNN | Word | 0.001 | 128 | 25 | 0.36 | 29.75 |
| Multi-RNN | Word2vec | 0.0001 | 128 | 20 | 0.19 | 15.67 |
| Bi-RNN | Word2vec | 0.001 | 128 | 15 | 0.66 | 8.08 |

**Top choices for model building**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Network** | **Embedding** | **Learning Rate** | **Batch Size** | **Epochs** |
| RNN | Word | 0.01 | 256 | 20 |
| Bi-RNN | Word | 0.01 | 128 | 30 |
| Bi-RNN | Word2vec | 0.001 | 128 | 15 |

**Note**:

* Character wise embeddings were not generating sentences with proper spellings (So, not included in the table)
* Highlighted row in ‘green’ background was built after data pre-processing
* Only Word2Vec embedding with Bi-directional network structure was built with and without data pre-processing (Rest of them are built without pre-processing)