International Institute of Information technology – Hyderabad

CSE471 - Statistical Methods in A.I

Course Project report

Project ID – 11

Predict movie tagline from text synopsis

Project Guide: Satyam Mittal (satyam.mittal@students.iiit.ac.in)

Project Members:

1. Jyothsna – 2018900063 (jyothsna.munipalle@students.iiit.ac.in)
2. Surekha – 2018900085 (surekha.medapati@students.iiit.ac.in)
3. Madan – 2018900075 (madan.nandiwada@students.iiit.ac.in)
4. Karthikeyan – 2018900074 (Karthikeyan.arumugam@students.iiit.ac.in)

Table of Contents

Abstract :

Keywords : NLP, TensorFlow , Recurring neural networks, Sequence to Sequence , Number Batch, Python

Tools Used

Programming language: Python 3.X

Dev tools: Juypter notebooks, Colab.research.google.com

Toolkits Used: Tensorflow, Numpy, Pandas,

GIT Hub Repository URL: <https://github.com/karumugamio/SMAIProject-Team31>

Google Drive URL: <http://tiny.cc/team31gdriverepository>

Problem statement:

For any given movie synopsis as a text content , is it possible to predict the movie tagline. ?

For Example : If Synopsis of Baahubali movie is given is it possible to predict tagline such as “The Beginning”

In other words, This problem statement is text abstraction problem. We can use NLP Technics to train a recurring neural network.

Literature review:

We found three journals related to similar problem statement.

Our work is inspired by approaches followed by above mentioned journals.

Challenges:

Data analysis and Clean up:

Data source: Kaggle Link

Issues with Data:

1. Missing values
2. Illegal characters and Symbols
3. Non-English Characters

Solutions and workaround:

Implementation:

Bi Directional LSTM – RNN.

Batching

Stop Criteria

Performance metrics

Rogue Score. – Theory and details

\*\*\* Put validation metrics as table+ Data Set details

Results ? Details ref the paper for the format

Conclusion :

References :

1. Predicting Movie Genres Based on Plot Summaries: https://arxiv.org/pdf/1801.04813.pdf
2. Folksonomication: Predicting Tags for Movies from Plot Synopses Using Emotion Flow Encoded Neural Network - https://aclweb.org/anthology/C18-1244
3. Patent Abstract Summarization using Recurrent Neural Networks – Reference Link Missing
4. Data Set Source – Kaggle Dataset for Movie Data set – Link Here:
5. Number batch – Details link
6. Number batch GIT Hub Link -
7. Glove - https://nlp.stanford.edu/projects/glove/
8. Rouge scores : https://stats.stackexchange.com/questions/301626/interpreting-rouge-scores
9. RNN – Sequence to Sequence Model : https://towardsdatascience.com/seq2seq-model-in-tensorflow-ec0c557e560f