Project Proposal

Semi-supervised Sequence Learning

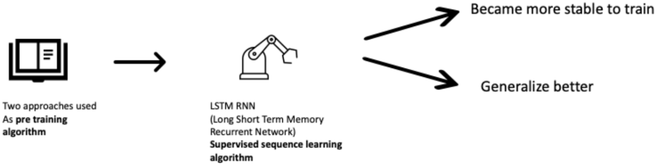
Sabiha Sultana[1], Piyakorn Munegan[2],
Sanjay Kanakkot Viswanathan[3],
Mohammed Rizwan Amanullah[4]

Department of Computing,
Macquarie University

October 12, 2021

Overview

- In this paper, they have introduced two approaches, to use unlabelled data to improve sequence learning with Recurrent network
  - **First approach (LM-LSTM)** Language modelling. Predict what comes next in a sequence
  - **Second approach (SA-LSTM)** Sequence auto encoder. Reads input sequence into vectors and predict the input sequence again.



Two approaches used
As **pre training algorithm**

LSTM RNN
(Long Short Term Memory Recurrent Network)
Supervised sequence learning algorithm

Became more stable to train

Generalize better

Claims - Pre-trained LSTM Results in strong performance in many classification tasks.
- Pre-trained LSTM performs better than LSTM initialized randomly.
- It is possible to use unsupervised learning with more unlabelled data to improve supervised learning.

Justification for Project Basis

- In Google Scholar this research paper has been cited by 989 people.
- Research paper has been published at Neural Information Processing Systems (NIPS 2015), CORE ranking A*. Link to Paper.
- Code has been made available on the official Tensor Flow GitHub Page.
- It covers in the area of Natural Language Processing(NPL); Recurrent Neural Networks (RNNs), Long Short-Term Memory recurrent networks (LSTM RNNs) and Sequence Autoencoder Long Short-Term Memory recurrent networks (SA-LSTMs).

Replication of original work

In this paper there were 4 data set IMDb, Rotten Tomatoes, DBpedia, 20 newsgroups and different technique used for the same.

Table: A summary of the error rates of SA-LSTMs and previous best reported results.

Dataset	**SA-LSTM**	**Previous best results**
IMDB	7.24%	7.42%
Rotten Tomatoes	16.7%	18.5%
20 Newsgroups	15.6%	17.1%
DBpedia	1.19%	1.74%

New Data Generation

The Process

1. Creating Python Script
2. Scrapping the data
3. Eg :Using "BeautifulSoup" python package

Using Python Script we will scrap the recent reviews replicate the data similar way with other Wikipedia, Rotten Tomatoes, GMB, DBpedia, Apple Reviews data set etc.

The End