ICAT Day: Neural Network Doc Summarization

CS4624 Multimedia, Hypertext, and Information Access

Team: Junjie Cheng

Instructor: Dr. Edward A. Fox

Virginia Tech, Blacksburg VA 24061, Apr 30th, 2018

Outline

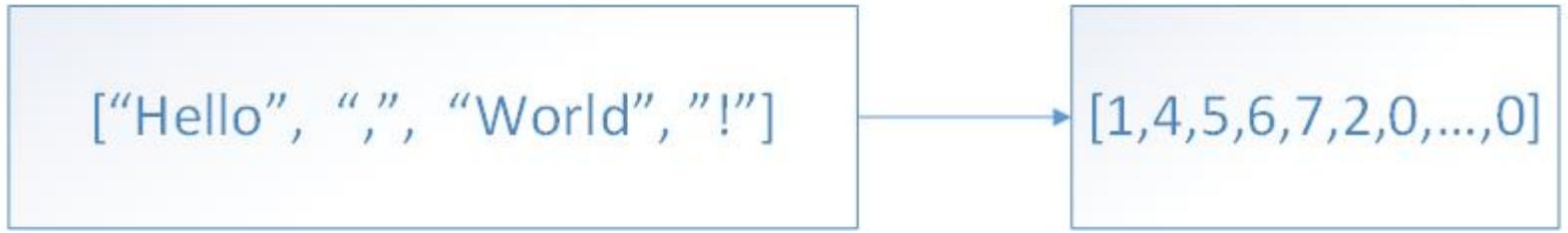
- Project Overview
- Data Preprocess
- Model Architecture
- Training
- Model Performance
- References and Acknowledgements

Project Overview

- Purpose: generate summarization from long document through deep learning.
- Model: sequence to sequence model with RNN.
- Dataset: CNN/Daily Mail news.

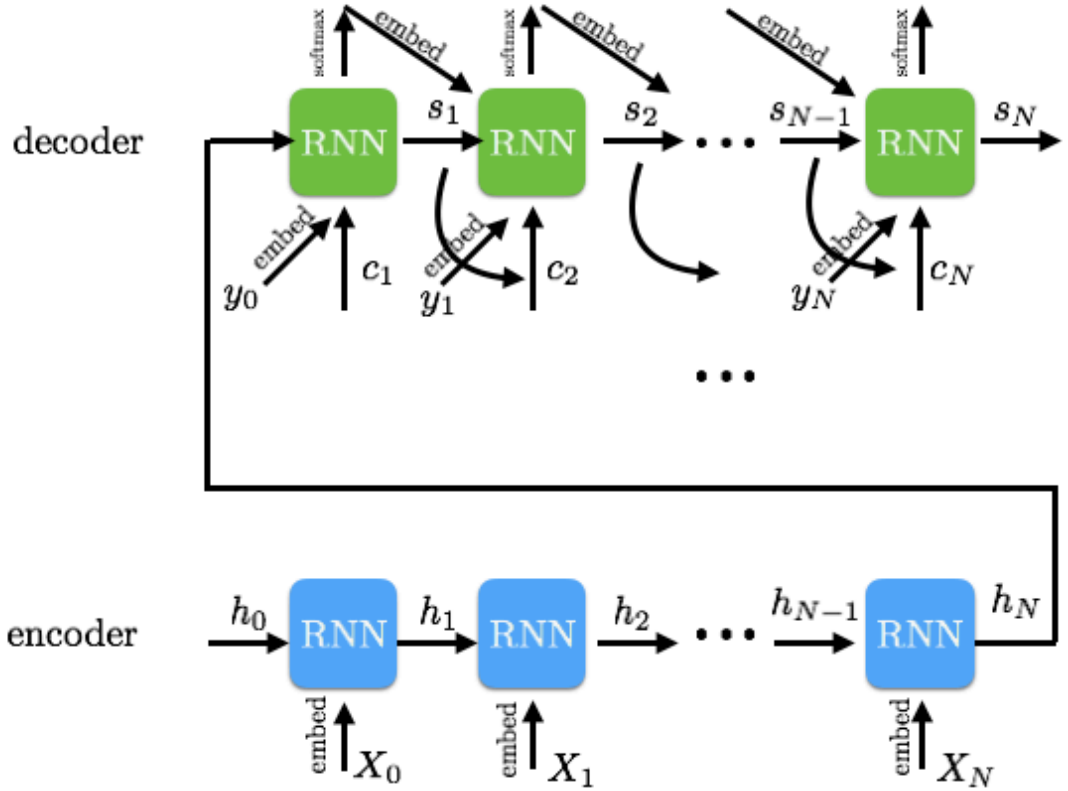
Data Preprocess

- Vocab size: 50000
- Input sequence max length: 400
- Target sequence max length: 100

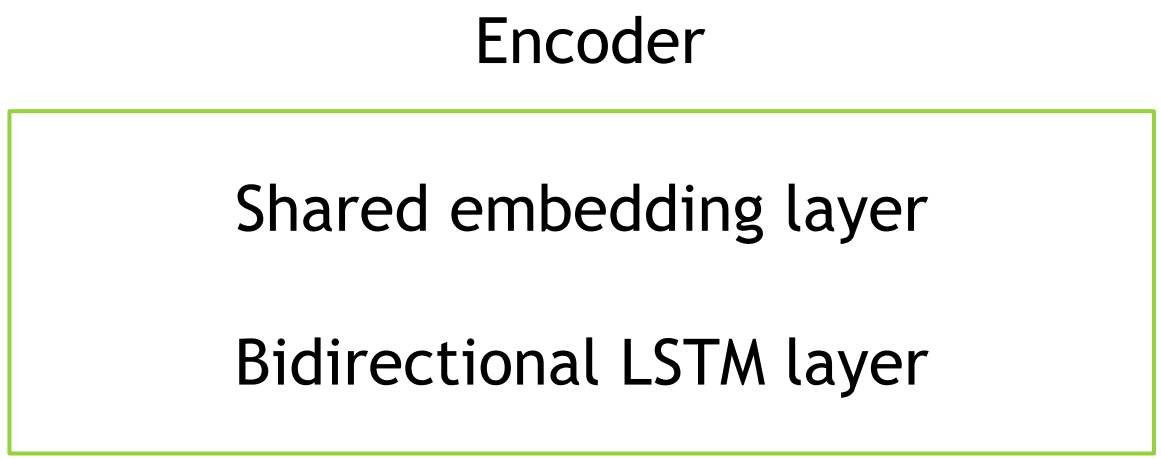


Model Architecture

Sequence to Sequence Model



Encoder Architecture



Encoder Workflow

Embedding layer

- Embedded Input sequence

LSTM layer

- Context
- Last hidden vector
- Last LSTM cell state

Decoder Architecture

Decoder

Shared embedding layer

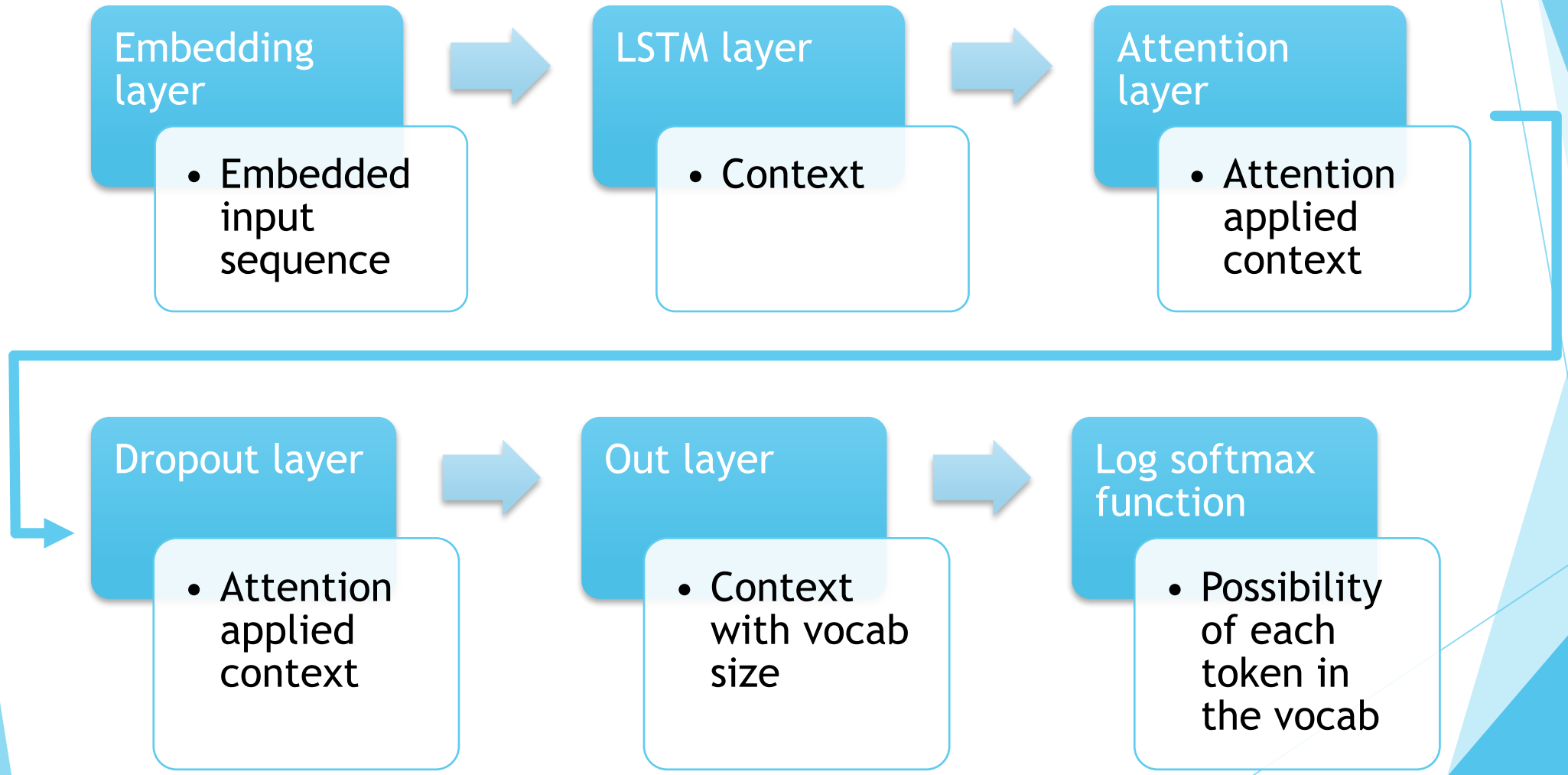
LSTM layer

MLP attention Layer

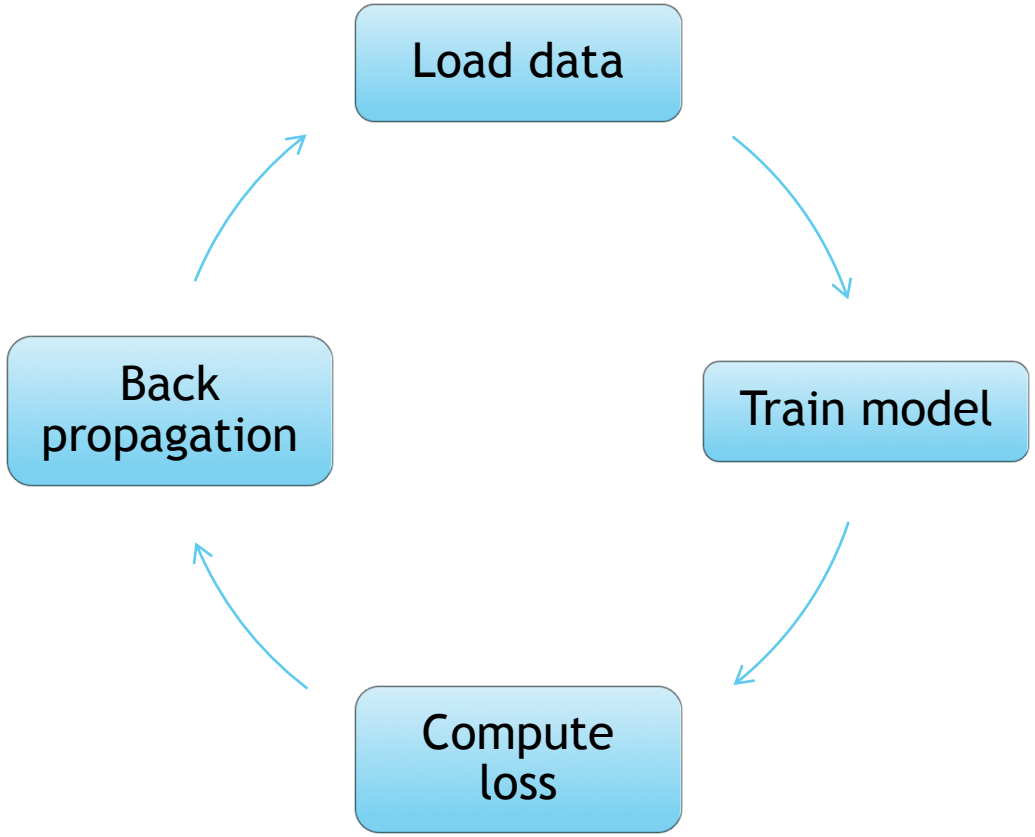
Dropout layer

Out layer

Decoder Workflow



Training Workflow



Training Architecture

- Optimizer: SGD
- Criterion: NLLLoss

- Batch size: 3
- Epoch number: 100
- Loss: 6.7 → 1.4
- Learning rate: 1
- Hidden size: 256
- Word embedding size: 128

Model Performance

- Demo

Acknowledgements

- Client: Yufeng Ma

Reference

- Figure Encoder-Decoder: Encoder-decoder: https://theneuralperspective.com/2016/11/20/recurrent- neural-networks-rnn-part-3-encoder-decoder/