Final Presentation: 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

[“Hello”, “,”, “World”, ”!”] → [1,4,5,6,7,2,0,...,0]
Model Architecture

Sequence to Sequence Model

[Diagram showing RNNs for encoder and decoder with embedding layers and sequence inputs and outputs]
Encoder Architecture

Encoder

Shared embedding layer

Bidirectional LSTM layer
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

- **Embedding layer**
  - Embedded input sequence

- **LSTM layer**
  - Context

- **Attention layer**
  - Attention applied context

- **Dropout layer**
  - Attention applied context

- **Out layer**
  - Context with vocab size

- **Log softmax function**
  - Possibility of each token in the vocab
Training Workflow

1. Load data
2. Train model
3. Compute loss
4. Backpropagation
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

- Generated summary: “have beaten three of their last three league games. the <UNK> scored in the second half of the last minute. the win takes all three points to move ahead of champions league place”

- Human-produced summary: “two goals from lionel messi help barcelona to a 3-1 win over almeria. kaka bags brace as real madrid coast to 3-0 victory at athletic bilbao. inter milan move up to second place in serie a with 2-0 win over chievo.”
Acknowledgements

- Client: Yufeng Ma
- Mr. Ma is a PhD student at Virginia Tech. He worked as the client of this project and guided the project through all project phases.