Assignment-2 Discussion (Question Answering)

Divyank Pratap Tiwari, 213050029
Singamsetty Sandeep, 213050064
Sakharam Sahadeo Gawade, 213050027

09 March 2022

Problem Statement (1/2)

- Create a question answering system using a Recurrent Neural Network (RNN)
- Dataset: SQUAD (https://rajpurkar.github.io/SQuAD-explorer/).
- Consider only FACTOID type questions. Answers to such questions are single words and phrases. Thus, only WHO, WHAT, WHEN, WHERE questions are admitted and not WHY and HOW (pl do not quibble about 'what is the reason of' and 'what is the process of' which are equivalent to 'why' and 'how'!).
- Create an encoder decoder network for the above task.

Problem Statement (2/2)

- Extract factoid QA pairs from SQUAD.
- Train with this data
- QA system performance metrics: precision, recall and F-score.
- Compare numbers with the state of the art. SQUAD is the standard benchmark for QA and umpteen number of papers report results on this data. A part of the assignment will be to tabulate these reported results along with their corresponding techniques.

Dataset Discussion

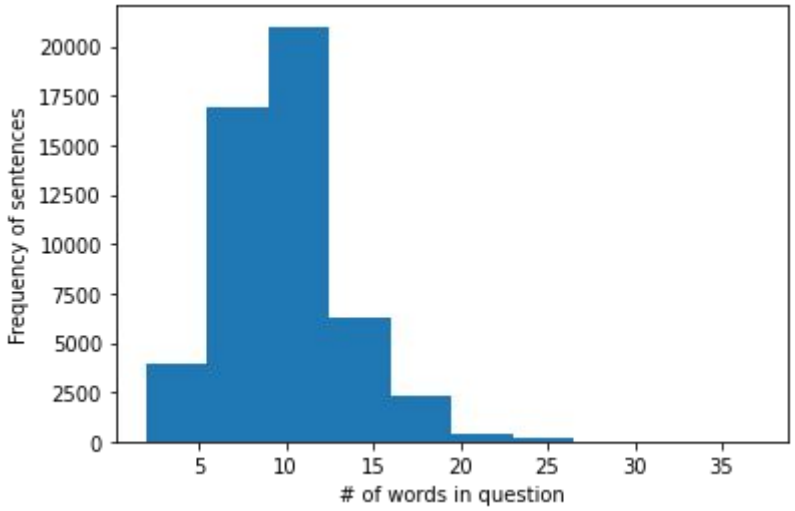
Details of Questions

- Show distribution of question: who, what, when, where

  Total Questions: 130319 Factoid questions: 51033

  What Questions: 41735 When Questions: 5455

  Where Questions: 3344 Who Questions: 8090



System implementation

Details of the RNN N/W

- Layers

  Encoder: One hot input → LSTM

  Decoder: LSTM→ One hot output
- Neurons/layer: 1 LSTM layer
- Embedding dimension Latent dimension: 256
- Attention: Not used
- Different Hyper parameters
  - Model learning rate: 1e-3
  - Adam learning rate: 1e-3
  - Batch size: 64

Various performance parameters: P, R, F-score, Accuracy

- PRF overall

  Precision: 0.008566666666666667

  Recall: 0.009216666666666665

  F1: 0.00852896825396825 4
- PRF for ‘who’, ‘what’, ‘when’, ‘where’ separately

	precision	recall	F1
who	0.011217	0.009935	0.00979
what	0.005921	0.006403	0.005971
when	0.0053191	0.01063	0.007092
where	0.033730	0.0317460	0.0321995

Error Analysis

- Good cases

  Question: What architecture type came before Norman in England?

  Predicted Answer: Anglo Saxon

  Actual Answer: Anglo Saxon

  Question: In what century was the Olympic balancing system used?

  Predicted Answer: 19th

  Actual Answer: 19th

- Bad cases

  Question: What is the population of the second largest city in California?

  Predicted Answer: 12 inches

  Actual Answer: 1.3 million

  Question: What percentage of farmland grows wheat?

  Predicted Answer: 7%

  Actual Answer: 50%

Error Analysis

In the second example, in bad cases, we can observe, that the answer is of type percentage, and the model was able to predict a percentage.

Take away : **Model was able to classify the type of answer correctly.**

Few other cases:

- When the actual output is of type **year**, predicted output is of type **year**
- When the actual output is of type **Proper Noun**, predicted output is of type **Proper Noun**

Examples:

- **Question**: Who ruled the duchy of Normandy

  Predicted answer: Alexander II

  Actual answer: Richard I
- **Question**: When did Great Britain claim Australia?

  Predicted answer: 1786

  Actual answer: 1788

Thank You