Machine Reading Comprehension: Question and Answers System for News Domain

Sandipan Basu, Aravind Gaddala, Garima Tiwari, Pooja Chetan

greatlearning
Learning for Life

TEXAS McCombs
The University of Texas at Austin
McCombs School of Business

ABSTRACT

Machine reading comprehension aims to teach machines to understand a text like a human. Reading comprehension is an AI-complete task, which requires the system to process a piece of text, comprehend and be able to extract the span of text which is the answer to the user query.

This project attempts to build a Question Answering system(QnA System) in the News Domain, where Passages(Context) will be News articles, and anyone can ask a Question against it.
For training our model, we have used the Stanford Question and Answer (SQuAD) dataset. The key components of our model architecture are- Embedding Layer, RNN Layer, and the Attention Layer. We trained our models using a custom embedding layer and also using transfer learning from GloVe and the Universal Sentence Encoder. For the RNN Layer, we built variations of the RNN Layer - Vanilla LSTM Layer, Bidirectional LSTM, Stacked LSTM Layers along with bi-directional Attention Layer. BERT has inspired many recent NLP architectures, training approaches, and language models. Hence, we leveraged transfer learning and built a Transformer based model using BERT.
For all the Models we evaluated the Model performance using two metrics- F1 Score and Exact Match. We concluded that the BERT model is superior in all aspects of answering various types of Questions.
We think the model we have built has a wide application for Answering Questions from News articles and has a wide application not limiting to online or offline need literacy. This system can become a foundation technology on which a voice-based question and answers can be delivered. This model can be extended to multilingual news question and answering system
Despite constant advances and seemingly super-human performance on constrained domains, reasoning with a model's interpretability is very limited. For our future research, we would like to look at a few aspects to build this interpretation of model inference.

INTRODUCTION

Problem of identifying and extracting an answer from a paragraph can be solved by training Deep neural networks along with using word understanding using Transfer Learning. There has been tremendous progress in the field of NLP after the introduction of the BERT framework by Google. This achieved State-of-the-Art results on 11 individual NLP tasks. BERT has inspired many recent NLP architectures, training approaches, and language models, such as Google's TransformerXL, OpenAI's GPT-2, etc.
Proposed project attempt to build a Question Answering system(QnA System) in the News Domain, where Passages(Context) will be News articles. Proposed system uses LSTM, Transformer architecture like BERT, Embedding layers to achieve state of the art language model to extract answers. The current system can be extended to cater to multiple languages.

METHODS AND MATERIALS

These architecture components which we have considered to solve our problem are:
Data Preparation & Augmentation- Training Data
For training and validation our model, we have used the Stanford Question and Answer (SQUAD) dataset.
Embedding Layer - A layer in the model which can use transfer learning. We used Custom Embedding Layer with 100 Dimensions, GloVec Embedding with 300 Dimensions and Universal Sentence Encoder for English with 512 Dimensions
RNN - Model architecture based on recurrence neural network. We customised 4 variations of this. We used LSTM as our base architecture.
Attention Layer - Attentions are an improvement over Recurrent NN. We implemented 2 types of model on Attention
Transformer Architecture - A network architecture which primarily uses Attention. A BERT based model was developed which ended up been superior than others.
Bilinear Transformation: A custom network layer operation to capture the similarity between each Context Token and Question and compute the probabilities of each token being Start and End.
Custom Loss function : Custom loss computes the difference between the Predicted(Answer Start Token, Answer End Token) and True(Answer Start Token, Answer End Token)

Figure 1. Data set

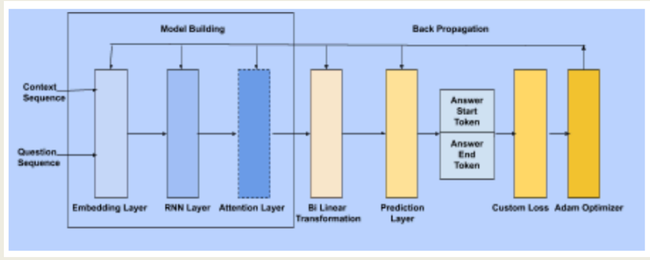

Figure 2. Model Arch

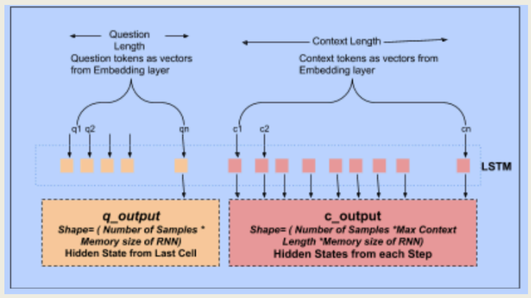

Figure 3. LSTM

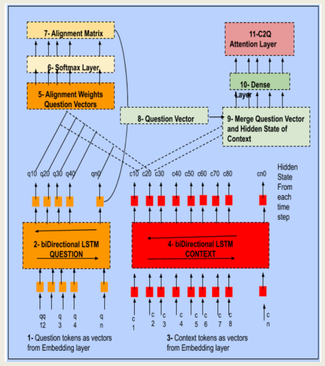

Figure 4. Attention

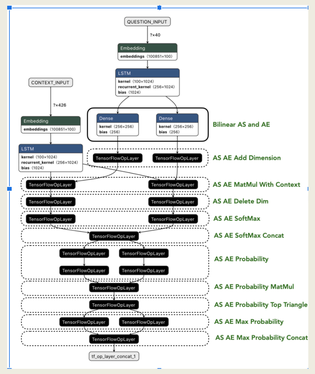

Figure 5. Model Arch 2

RESULTS

Original Data Set	130306 samples
Train set	78183 samples
Eval dataset	26061 samples
Test dataset	26062 samples

Figure 6. Dataset

	Model Name	F1(Ans)	EM(Ans)	F1(Plau Ans)	EM(Plau Ans)
0	LSTM Baseline	29.00112098	28.22500192	0.2146227416	0.05371805694
1	Deep LSTM + GloVe	26.66643901	26.03023559	0.2763011224	0.06906607321
2	Bi-LSTM + GloVe	28.02196074	27.52666718	0.2811061296	0.1266211342
3	Bi-LSTM + GloVe + C2Q Attention	33.09511933	33.09416008	0	0
4	Bi-LSTM + GloVe + Q2C-C2Q Attention	29.62215708	29.13053488	0.2000388679	0.06906607321
5	LSTM Baseline + Universal Sentence Encode	24.54525456	22.95295833	0.6185752402	0.09976210575
6	Bi-LSTM + Universal Sentence Encode	32.02522636	31.54401044	0.06841133647	0.0230220244
7	Bi-LSTM + C2Q Attention + Universal Sentence Encode	30.69597542	30.12048193	0.1469459993	0.04220704474
8	BERT + Cased_L-12_H-768_A-12 + DeepPavlov	59.09692802	51.36213644	22.53928947	18.17972527
9	BERT + Uncased_L-24_H-1024_A-24 + Huggingface	57.51315339	49.76977976	22.1581681	17.75381782

Table 7. Model Performance

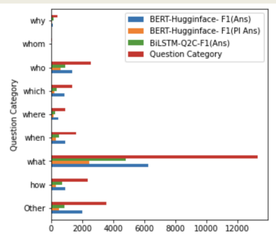

Table 8. Ques Category wise Performance

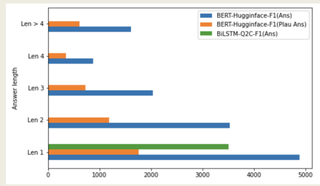

Table 8. Ans Length Performance

Best performing model is the model on BERT using transfer learning. We used pre-trained models from HuggingFace and DeepPavlov. Though this is the best model that we have, this required the least amount of effort on our part. For this project we concentrated our effort on building the RNN models with the Attention mechanism, which is explained in detail in this document.

DISCUSSION

This project aims to solve reading comprehension in the News domain, where Passages will be News articles, and anyone can ask a Question against it.

- News feeds be it online or offline needs literacy. This system can become a foundation technology on which a voice-based question and answers can be delivered. Huge implications in rural areas and specially in times where print media is unreachable
- Multilingual news question and answering system
- Help researchers who are mining news archives

The technology is very generic and can be applied to varied domain areas, namely -

- Help a financial consultant to ask a question given a document of his own and answer precisely as possible
- Help a teacher to check the answers of the student against a questionnaire in a specific subject
- Help to answer questions against a contract document
- Help build a knowledge base of products from a product's user manual and then provide quick answer for questions around it- It can be used to reply to user queries about it
- Application in legal tech- answering questions related to various law

BENCHMARK COMPARISON

	FI	EM
Human Performance Stanford University (Rajpurkar & Jia et al. '18)	89.452	86.831
SA-Net on Albert (ensemble) QIANXIN	93.011	90.724
LSTM Baseline Model	29	28
Bi-LSTM + GloVe + C2Q Attention	33.095119	33.094160
BERT + Uncased_L-24_H-1024_A-24 + Huggingface	57.513153 (+22.158168)[1]	49.769780 (+17.753818)[2]

CONCLUSIONS

This paper addresses the solution of answering extraction from a given context or paragraph. Various deep learning models in NLP have been discussed with primarily SQUAD dataset. CNN/DailyMail dataset was used to have a News domain specific information.
The overall goal of the proposed work is to determine the best performing architecture and build a production ready QnA AI system

REFERENCES

1. Pranav Rajpurkar.Robin Jia.Percy Liang. Know What You Don't Know: Unanswerable Questions for SQuAD. https://arxiv.org/pdf/1806.03822.pdf
2. Danqi Chen, Adam Fisch, Jason Weston & Antoine Bordes. Reading Wikipedia to Answer Open-Domain Questions.https://arxiv.org/pdf/1704.00051.pdf
3. Han Xiao - Engineering Lead @ Tencent AI Lab.Teach Machine to Comprehend Text and Answer Question withTensorflow. https://bit.ly/2WUQdje
4. Jay Allamar. Visualizing A Neural Machine Translation Model (Mechanics of Seq2seq Models With Attention). https://bit.ly/3eWolBi
5. Daily Mail Reading Comprehension Task.https://arxiv.org/pdf/1606.02858v2.pdf