Question Answering System

-Sneha Kotha
B00839933

INTRODUCTION

Problem statement: Given a paragraph and a unambiguous question about the paragraph, the system is trained to answer the question with appropriate solution.

Dataset : SQuAD
Training : dev data 8:1 ratio
The remaining is used as the testing set.

Many of the successful model till now have a common structure based on layers:

- -Embedding layer : word embeddings
- -Attention layer.
- -Modeling layer
- -Output layer

RELATED WORK

SQUAD let to significant amount of papers in the last two years[1].

- Wang & Jiang (2016b) - ***Answer Boundaries.***
- Seo et al. (2016) introduce ***Bi-directional attention flow networks***.
- Xiong et al. (2016) propose ***Dynamic co-attention networks.***
- RNET(2018) uses ***Gated attention*** between the question and context and ***Self attention layer***.

DATASET & EMBEDDING LAYERS

Features:

- length of paragraph, questions, answer span.
- And to choose the architecture

Example: 20 common words in question -paragraph we can use word embeddings using glove with 100 dimensions.

Also if there are too many word embeddings it causes **overfitting**.

Word embeddings : we are using pertained word vector (GLOVE[4])

Contextual embedding layer : Uses Long Short-Term Memory Network (LSTM) on top of the embeddings.

Attention Flow layer : This layer is important as to allows information from question and passage to be encapsulated within context word representation.

Types of Attention layers :

- BiDAF
- Contention
- Self attention

Combining self-attention and BiDaf[2] - F1 Score increased by 1%.

Modeling layer:

- Consists of **two layers of bi-directional LSTM**, with the output size of 'd' for each direction.
- BiDAF - approximately gives **2% raise** on the baseline.
- Prediction algorithm - increased the start and end probabilities of the answer.
- Range of the answer length-Increase the F1 score by 2%

Output Layer :

Provides an answer to the query. *M* is passed to another bidirectional LSTM layer and obtain *M2*. Then we use *M2* to obtain the probability distribution of the end index.

ARCHITECTURE

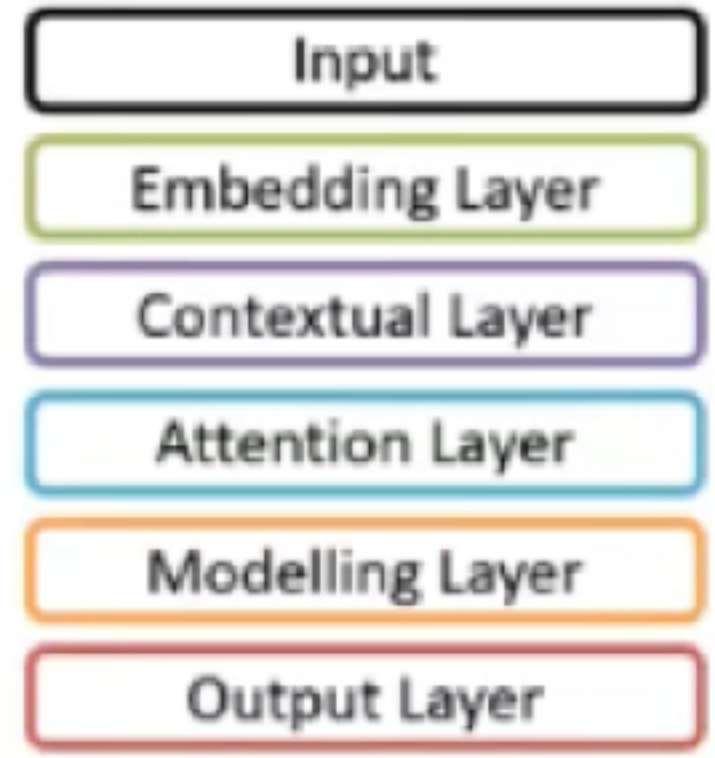
Input - Context and Question text

Embedding layer- Word embed

Contextual layer - bi directional LSTM and is concatenated.

Attention layer: - bi-attention feeds through-> self attention and is concatenated with the output.

And then finally to the modelling layer with 2 LSTMs



ARCHITECTURE

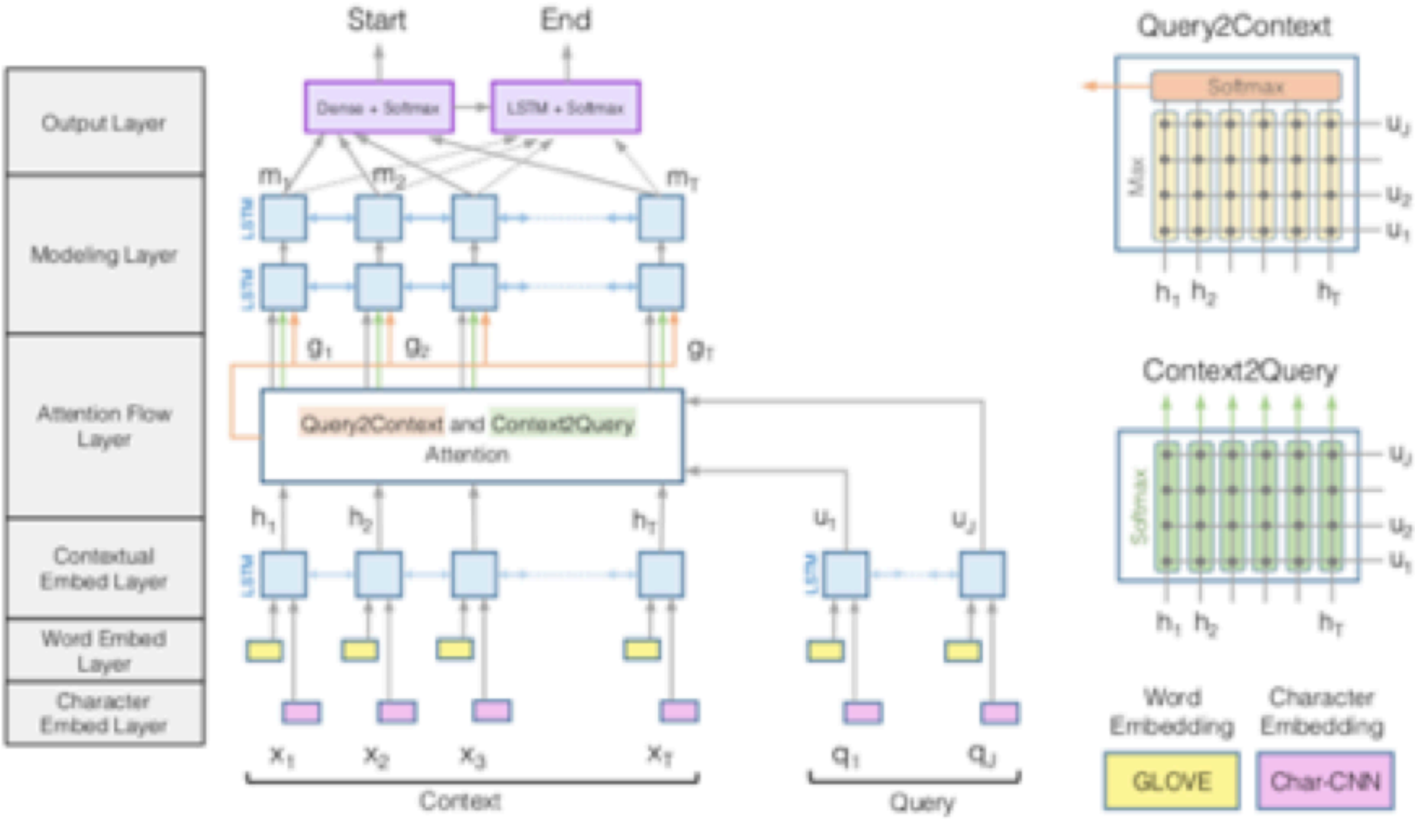


Figure 1: BiDirectional Attention Flow Model *(best viewed in color)*

EXPERIMENTS

Table 1: Leaderboard Performance

Model	Dev EM	Dev F1	Test EM	Test F1
BiDAF Attention + BidaF decoder (Ours)	64.427	74.444	65.593	75.142
BiDAF Attention + Answer Pointer decoder (Ours)	63.652	74.024	64.597	74.364
BiDAF Attention + Self Attention (Ours)	63.500	73.638	-	-
BiDAF Attention + Self Attention + Answer Pointer decoder (Ours)	63.954	74.051	-	-
Coattention Only (Ours)	53.955	64.171	-	-
Coattention with BidaF decoder (Ours)	58.079	69.008	-	-
Coattention with Answer Pointer decoder (Ours)	58.543	69.390	-	-
BiDAF	-	76.3	-	76.9
DCN	65.4	75.6	66.2	75.9
SQUAD Baseline(Ours)	34.853	44.117	-	-
Human (Rajpurkar et al., 2016)	81.4	91.0	82.3	91.2

Best score = 75%

RESULTS

Question Type	F1 (%)	EM (%)
When	82.95	76.56
Who	76.41	70.35
How	74.44	65.21
What	73.08	63.05
Where	72.83	62.11
Why	59.84	37.86

F1 AND EM SCORES

This model achieved 75.11% F1 and 64.75% EM on the dev set.

Where is it used?

- Legal documents
- Banking & Insurance
- Do we have any flight?
- Network issues

!?

FUTURE WORK

- Implementing **Character level CNN** can improve performance on out-of-vocabulary words.
- Using association of models and more **word features**.
- Dynamic co-attention Networks[3]
- Dynamic Pointer Decoders.

References:

[1] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, Squad: 100,000+ questions for machine comprehension of text, arXiv preprint arXiv:1606.05250, 2016.

[2] M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi, Bidirectional attention flow for machine comprehension, arXiv preprint arXiv:1611.01603, 2016.

[3] C. Xiong, V. Zhong, and R. Socher, Dynamic coattention networks for question answering, arXiv preprint arXiv:1611.01604, 2016.

[4] J. Pennington, R. Socher, and C. D. Manning, Glove: Global vectors for word representation., in EMNLP, vol. 14,pp. 15321543, 2014.

Any

Question ?? Answering

THANK YOU :)