PA – 4 Question Answer Task on Squad

By – Mudit Jain(muditjai@)

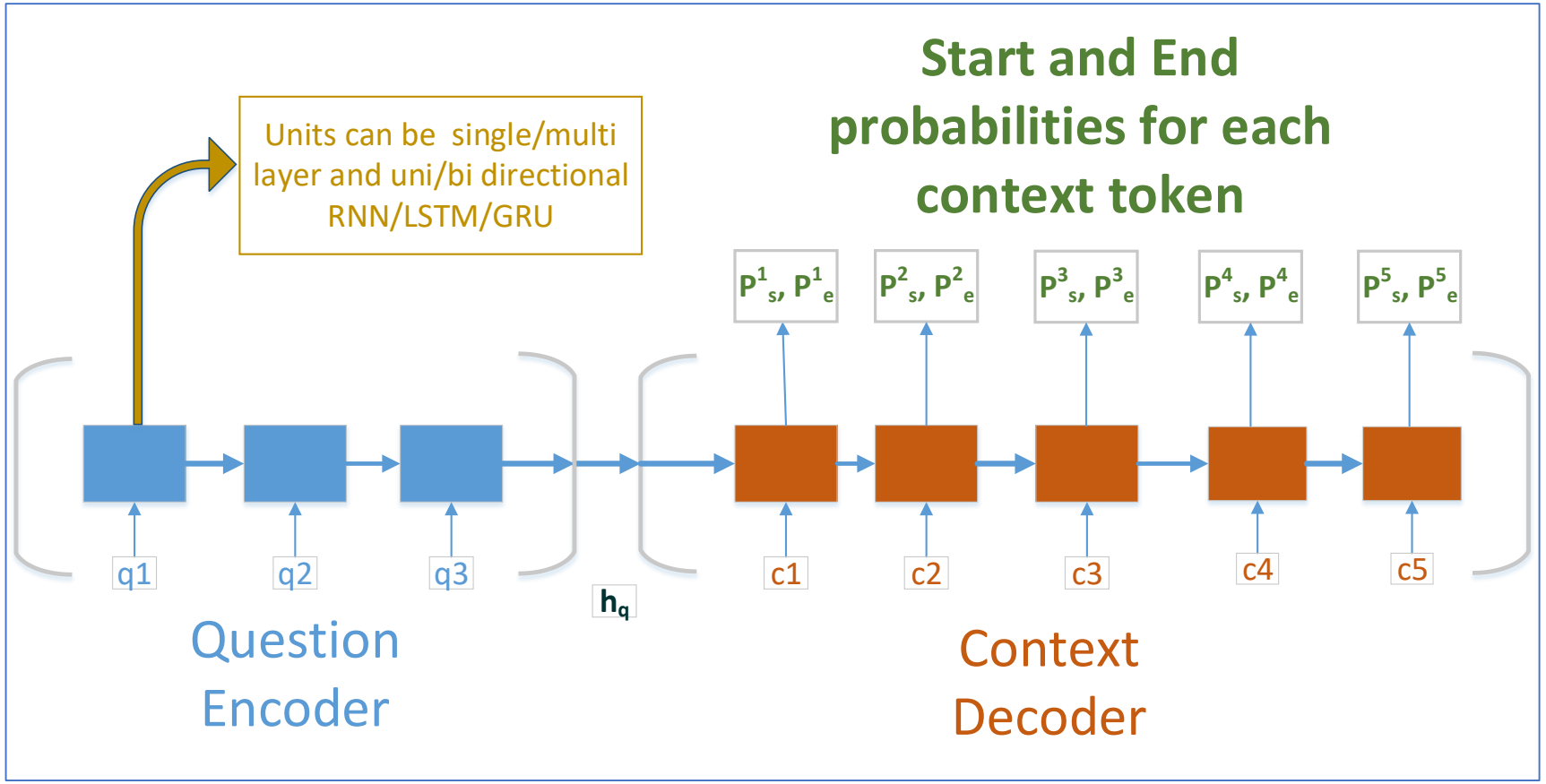
Introduction to problem and dataset

- Question answering is a challenging sub-problem within machine comprehension area.
- Solving question answering problems can help automate customer care, enable better search in text etc.
- Recently Squad dataset with 100K context, question, answers tuples has been created using Wikipedia and human judgment.
- The problem statement is – Given a context paragraph, and a question, generate an answer span pair $a_s < a_e$ which answers the question. The performance is measured using F1 score and Exact match(EM) score.

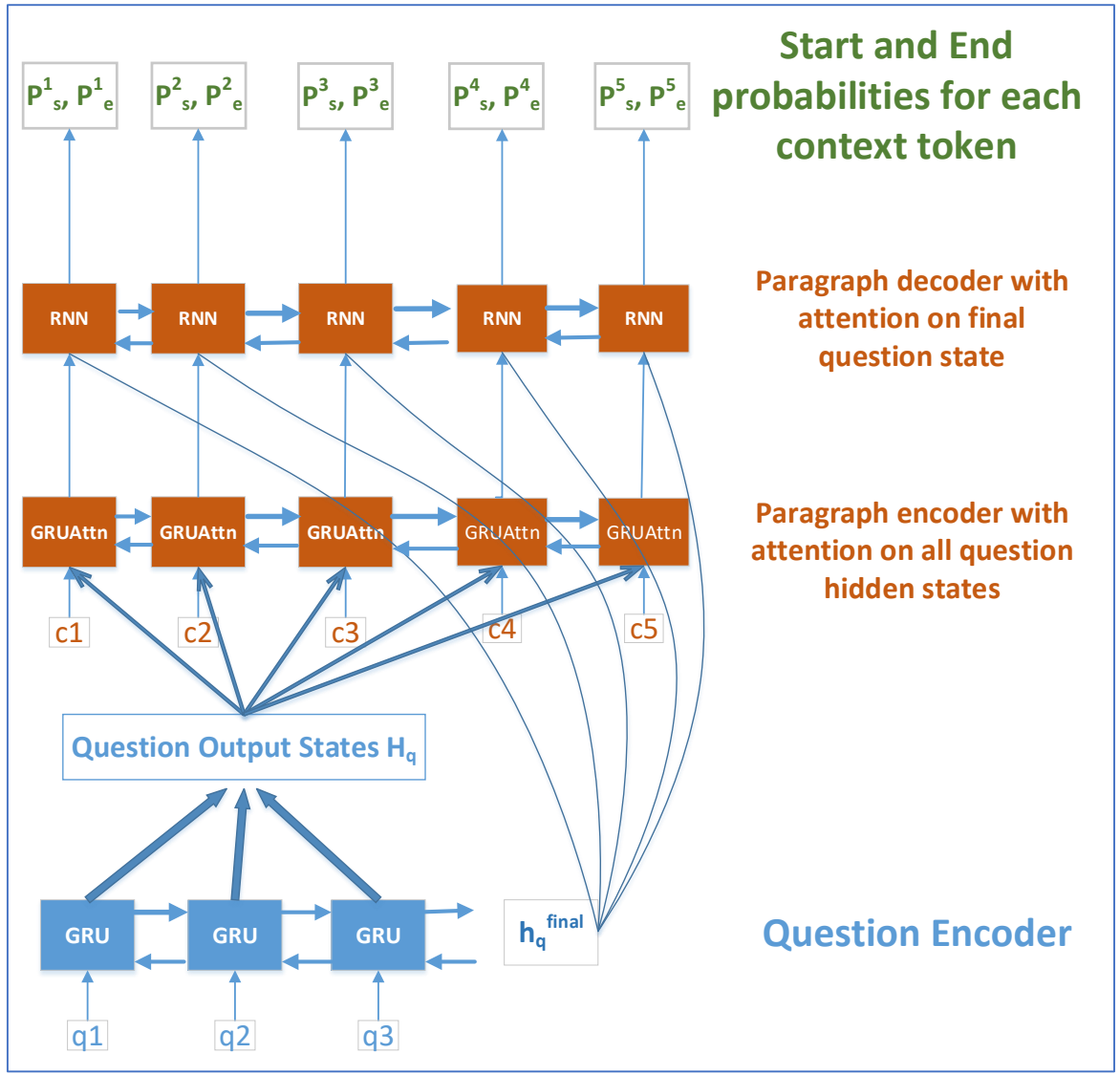
Solution approaches and conclusions

- Recent research has focused on using various deep learning methods to solve it.
- I have created 3 different NN model for this problem.
- Model 3 with 30/20 as F1/EM score, is able to learn the correct semantics of response ie “Who” type of questions need a person in response, “When” type of questions need a time, “How much” type of questions expect a quantity.
- Questions with simple sentence structures and short simple answers are answered correctly most of the times. Long answers are not answered correctly.
- Even when factually incorrect, the answers are still mostly coherent.

Model 1 – Simple encoder decoder network



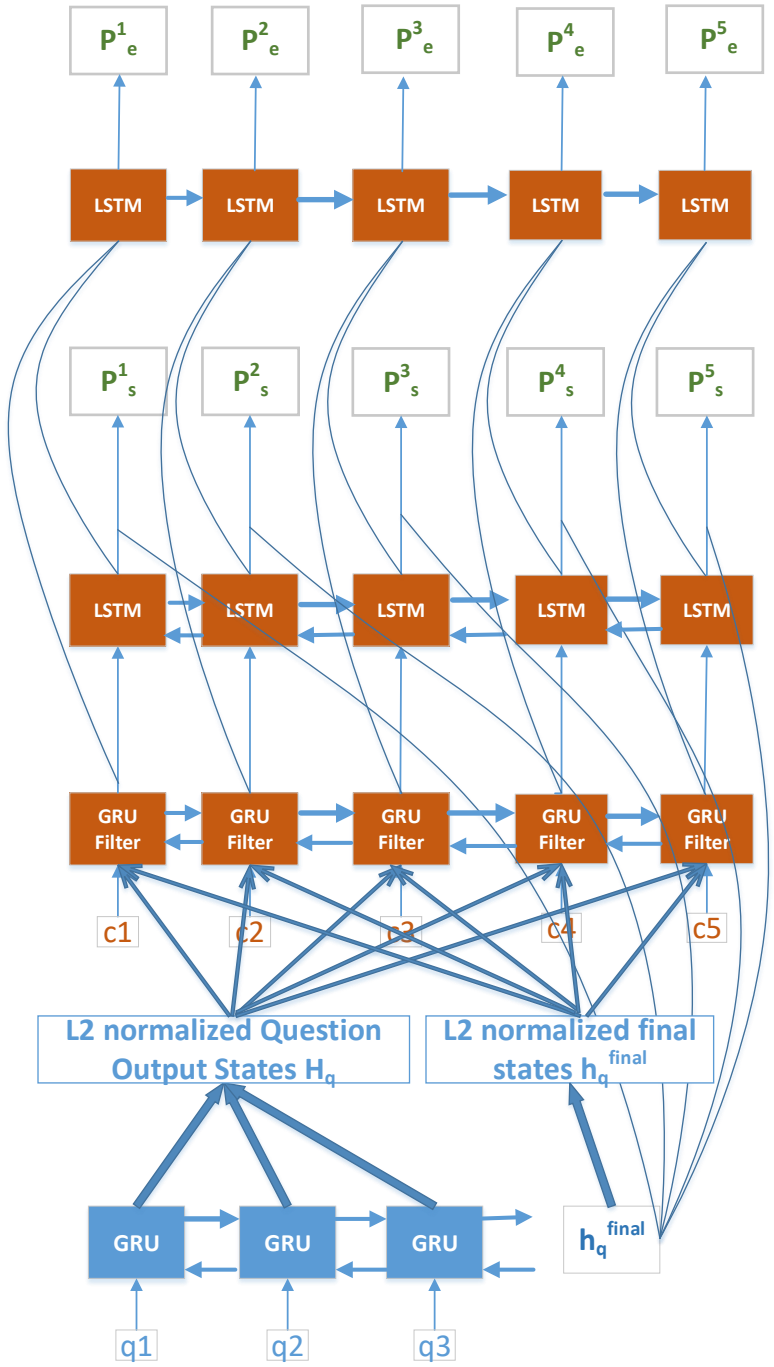
	Sanity Check Set		Dev Set		Test Set	
	F1	EM	F1	EM	F1	EM
Simple Encoder Network	3.34	0.37	5.20	0.64	4.89	0.63



Model 2 – Attention based encoder decoder network

	Sanity Check Set	
	F1	EM
Attention based model epoch 4	2.33	0.12
Attention based model epoch 8	2.27	0.12

Model 3 – Filter and Attention based two step decoder network



GRU Filter Cell

$$h_p = GRU(e_p)$$

$$\alpha = cosine(h_p, H_q)$$

$$h_p^{max\_sim} = h_p * max(\alpha)$$

$$h_p^{mean\_sim} = h_p * mean(\alpha)$$

$$h_p^{question\_sim} = h_p * cosine(h_p, h_q^{final})$$

$$gru\_out = Linear(h_p^{max\_sim} h_p^{mean\_sim} h_p^{question\_sim} h_q^{final})$$

Model 3 Performance

	Sanity Check Set		Dev Set		Test Set	
	F1	EM	F1	EM	F1	EM
GRU filter attn epoch 1	18.90	11.23	17.65	8.24	17.36	7.89
GRU filter attn epoch 2	25.88	18.39	24.39	13.3	--	--
GRU filter attn epoch 3	28.04	20.61	28.87	17.38	--	--

Thanks!