Building End-to-End QA pipeline with RL

ETI RASTOGI

PRASHANT BUDANIA

Overview

- **Objective**
- **Introduction**
- **Datasets**
- **BiDAF**
  - Architecture
  - Results
  - Error Analysis
  - Discussions
- **Reinforcement Learning**
  - REINFORCE algorithm
  - Results
  - Discussions
- **Conclusion**

Objective

Our objective for this course was two fold:

1. Implement a baseline neural network based model for question answering on datasets where the goal is to do span selection given a query and a snippet
2. For SearchQA dataset, implement a Reinforcement Learning base passage selection module which can help us select one snippet which then can be used with baseline QA model

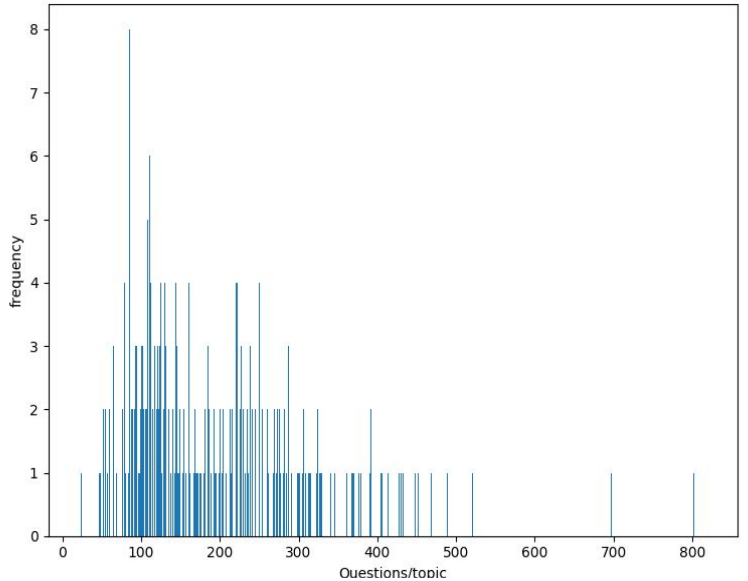
Introduction: Initial Proposal

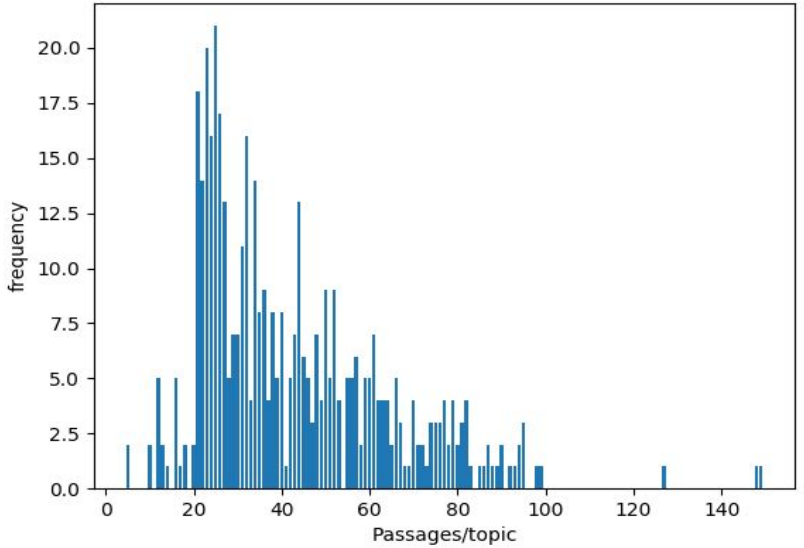
- Implement BiDAF model as baseline QA model
- Get results on SearchQA and WikiSuggest datasets and perform preliminary error analysis
- Add modules based on the initial results:
  - Passage ranking module for SearchQA dataset
- Implement the joint model (A Joint Model for Question Answering and Question Generation) (Wang et al 2017)
- If time permits:
  - Query reformulation – Sequence to Sequence translation (for queries) on SearchQA

Datasets

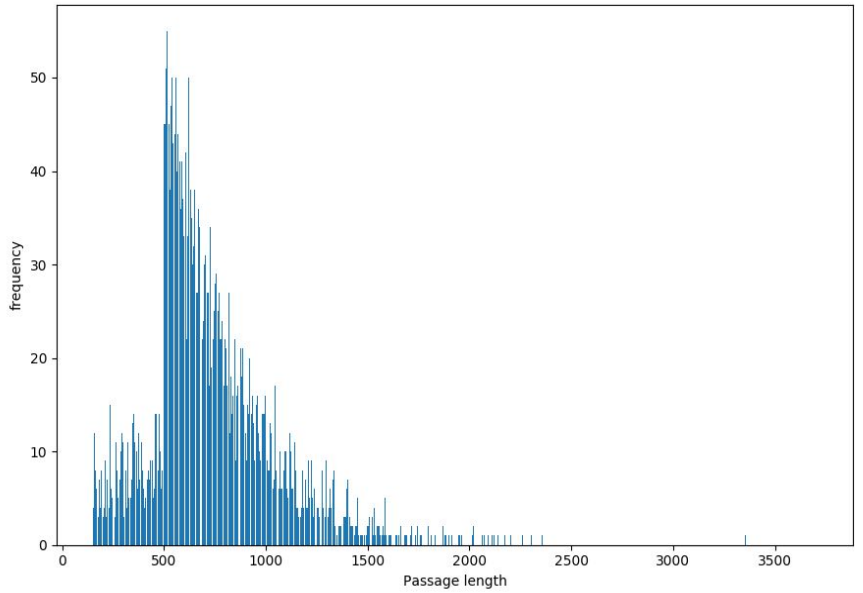
SQuAD(Stanford Question Answering Dataset)

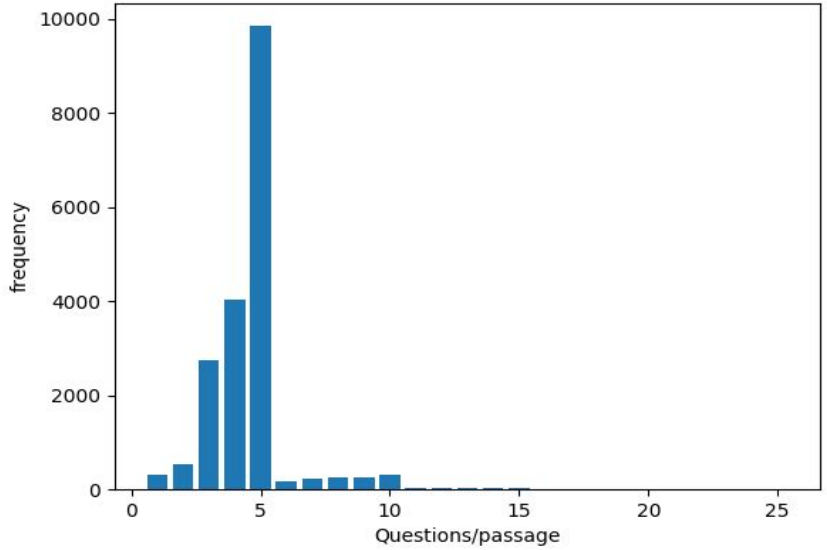
	Number of questions	Number of topics	Number of paragraphs
Train	87599	442	18896
Validation	10570	48	2067

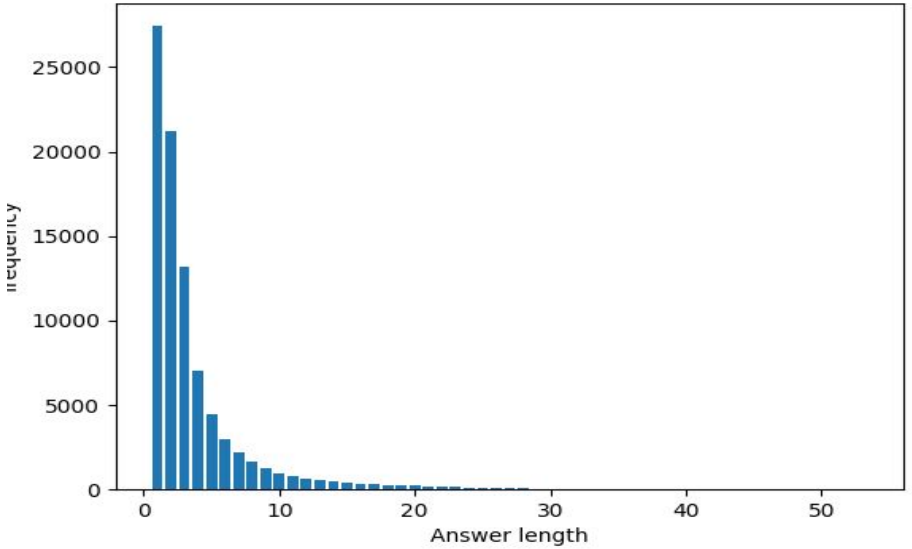




Average passage length	735.78
Average number of questions/passage	4.63
Average number of passages/topic	42.75
Average number of questions/topic	198.18
Average answer length	3.58



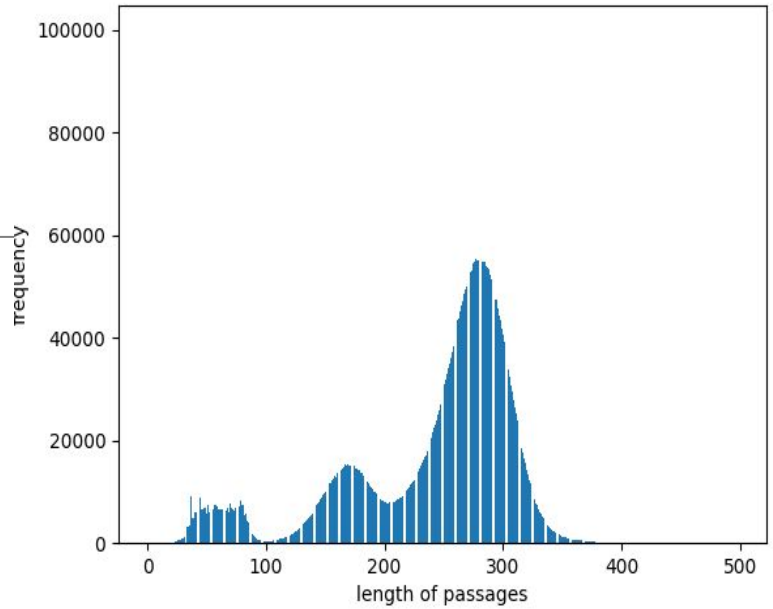


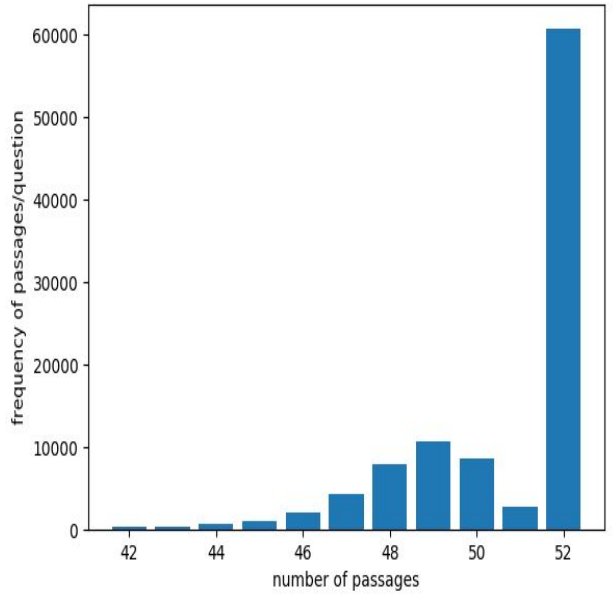


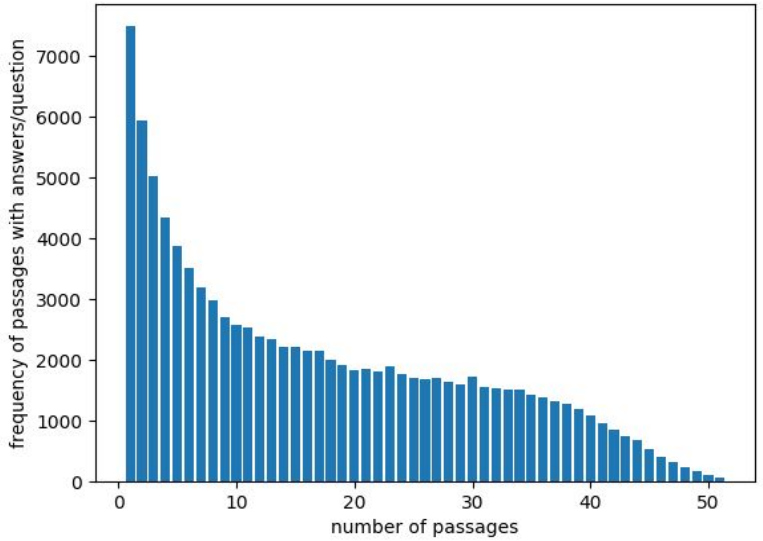
SearchQA

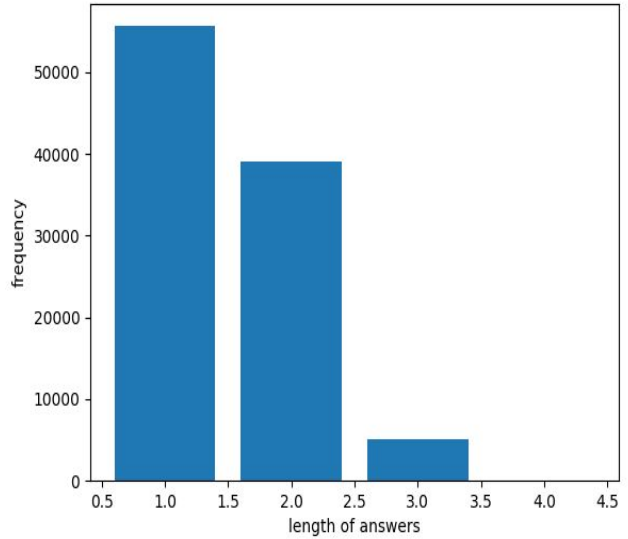
	Number of questions
Train	99820
Validation	13393
Test	27248

Average passage length	237
Average number of passages/question	50.6
Average number of passages with answers /question	16.78
Average answer length	1.5







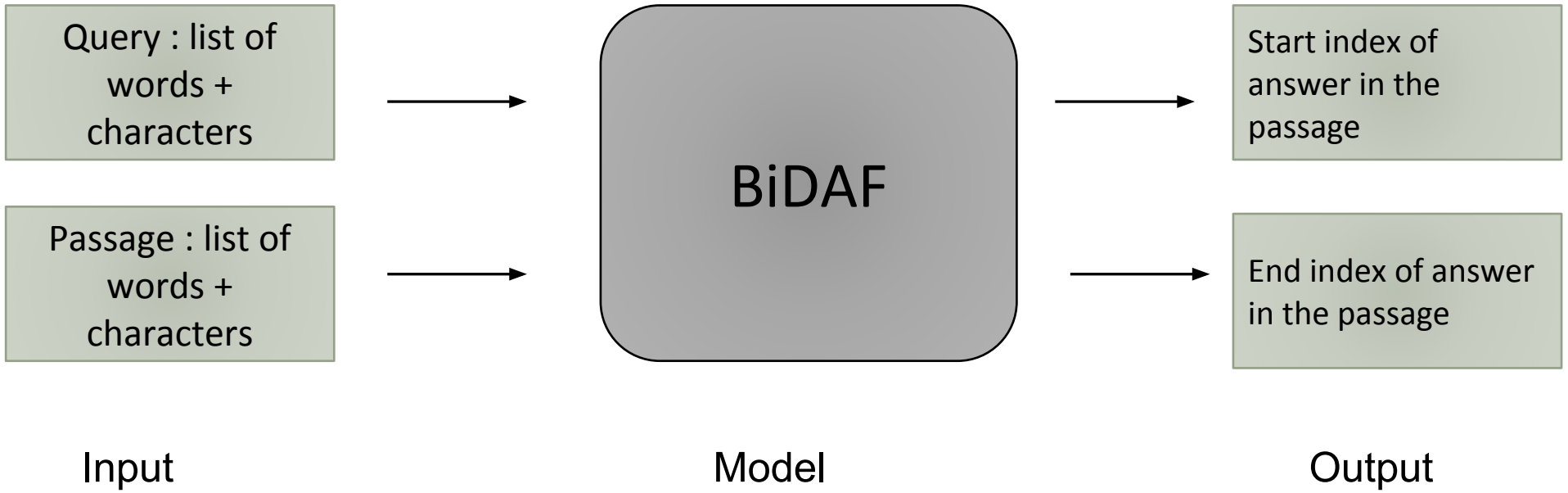


BiDAF : Bidirectional Attention Flow for Question Answering

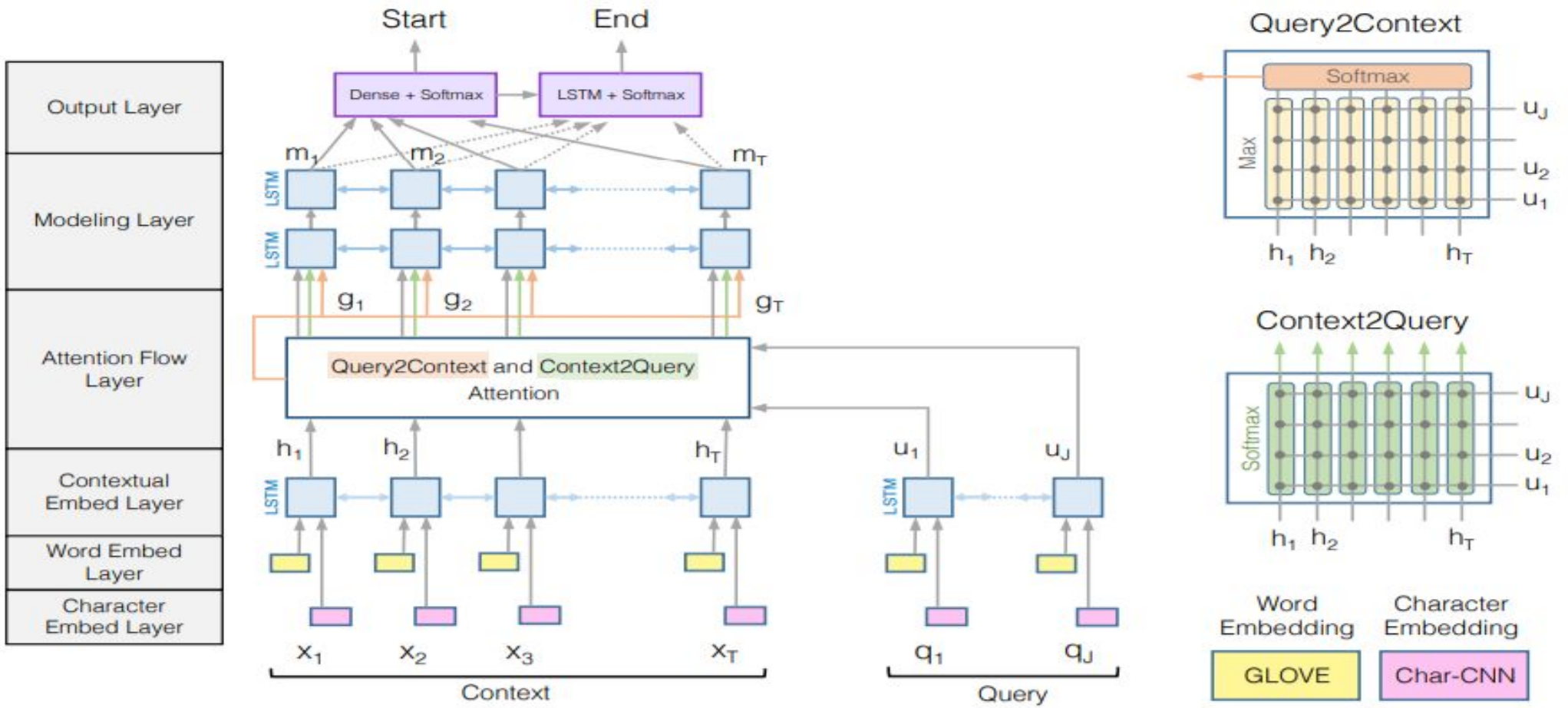
What is BiDAF and how is it different ?

- A hierarchical multi-stage architecture which models the paragraph representation at different levels of granularity.
- Performs attention in two directions to obtain a query-aware context representation.
- Reduces the information loss caused by early summarization.
- Uses a memory-less attention mechanism.

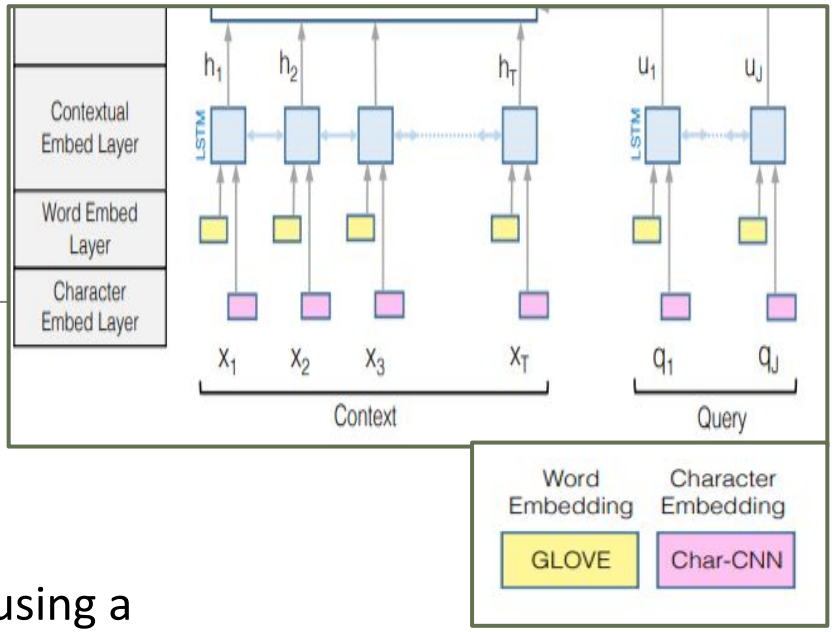
Architecture



Architecture



Architecture



1. **Character Embedding Layer** : each word to a vector space is mapped using character-level CNNs.

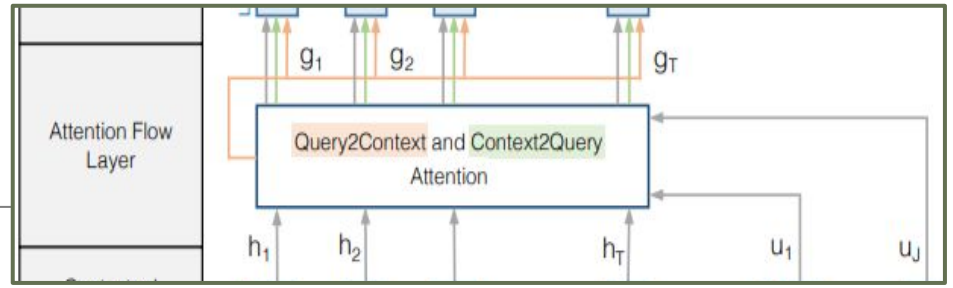
2. **Word Embedding Layer** : each word to a vector space is mapped using a pre-trained GLOVE embeddings

3. **Highway network** : concatenation of the character and word embedding vectors is passed to a two-layer Highway Network

$$\mathbf{y} = H(\mathbf{x}, \mathbf{W_H}) \cdot T(\mathbf{x}, \mathbf{W_T}) + \mathbf{x} \cdot (1 - T(\mathbf{x}, \mathbf{W_T})).$$

4. **Contextual Embedding Layer** : each embedding is further passed through a bidirectional LSTM to model the temporal interactions between words

Architecture



5. Attention FLow layer :

Similarity matrix : computes similarity between each context word with respect to every query word

$$S_{tj} = W^t[h_t; u_j; h_t \bullet u_j]$$

Context2Query : determines which query words are most relevant to each context word

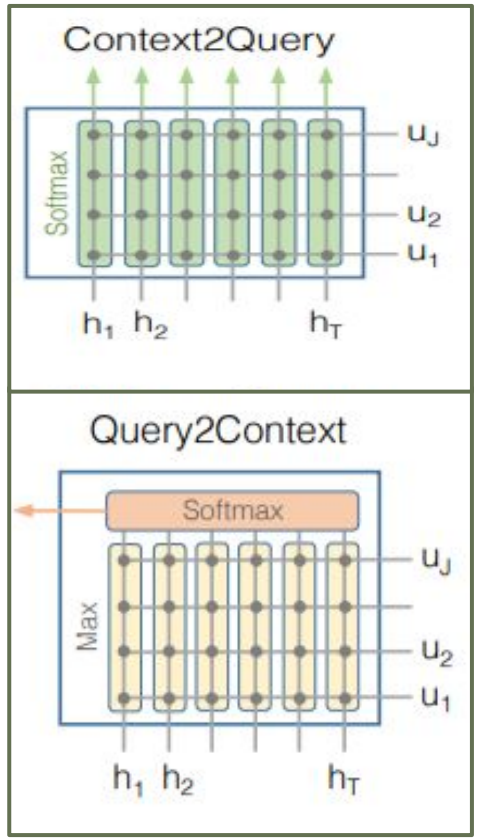
$$a_t = softmax(S_{:t})$$

$$c2q_t = \sum_j a_{tj} U_j$$

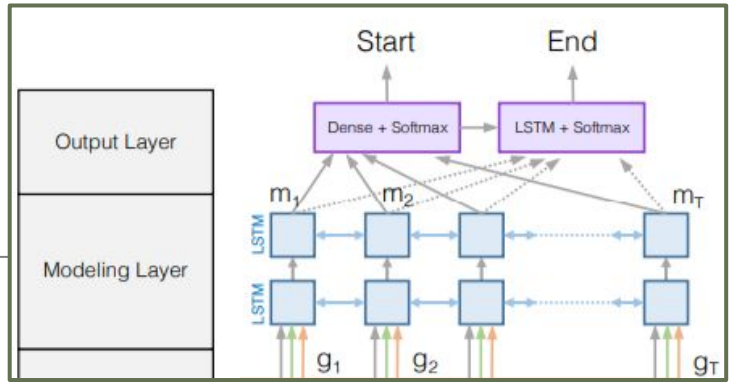
Query2Context : determines which context words have the closest similarity to one of the query words and are hence critical for answering the query.

$$b = softmax(max_{col} S)$$

$$q2c = \sum_t b_t H_t$$



Architecture



6. Modeling Layer :

Captures the interaction among the context words conditioned on the query.
Consists of 2 layer Bi-Directional LSTM input to which is given as :

$$G = [H; c2q; H \bullet c2q; H \bullet q2c]$$

7. Output Layer :

Predicts the start and the end indices of the phrase in the paragraph
Start index (p^1) :

$$\mathbf{p}^1 = \text{softmax}(\mathbf{w}_{(\mathbf{p}^1)}^\top [\mathbf{G}; \mathbf{M}])$$

End Index(p^2) : output is passed through a LSTM to obtain M^2 which is used to obtain the probability distribution of the end index

$$\mathbf{p}^2 = \text{softmax}(\mathbf{w}_{(\mathbf{p}^2)}^\top [\mathbf{G}; \mathbf{M}^2])$$

Training Loss

$$L(\theta) = -\frac{1}{N}\sum_{i}^{N} \log(\mathbf{p}^1_{y^1_i}) + \log(\mathbf{p}^2_{y^2_i})$$

Where N is the number of examples in the dataset, y^1_i and y^2_i are the true start and end indices of the i-th example.

Best Span Selection at Test Time

Use Dynamic Programming

The answer span (k, l) where k ≤ l is chosen as :

the maximum value of $p^1_k * p^2_l$

Results

SQuAD

	Accuracy(EM)(Ours)	Accuracy(EM)(Original)
With character embeddings	64.2	67.7
Without character embeddings	62.3	65.0
Reduced Vocabulary size	63.0	-

SearchQA

	Accuracy(EM)(Ours)
With character embeddings + Reduced Vocabulary size	27.8

Error Analysis : SQuAD

Question	Predicted	Gold Answer	Reason for Error	% error
Which articles of the Free Movement of Workers Regulation set out the primary provisions on equal treatment of workers?”	1 to 7	articles 1 to 7	-	53
What year did BSkyB acquire Sky Italia ?	2014	2014	Wrong span	12
when did French and Indian war ended ?	1754–1763	1763	Splitting on space	3

Error Analysis : SearchQA

Question	Predicted	Gold Answer	Reason for error
party u singer also plays young lady named hannah	hannah	Miley Cyrus	Wrong answer span First paragraph talks only about miley (no mention of hannah)
signature appetizer p f changs chicken cups vegetable	crisp lettuce	lettuce	-
4 x 12	4	48	Answer present but context is different.(passage talks about construction)
barbara undershaft		major barbara	Extra information required

Discussion

- Majority of errors in SQuAD are superficial(predicted answer is similar to golden answer)
- BiDAF does not perform well on SearchQA dataset. This might be because :
  - SearchQA is more complex
  - Concatenation of passages drastically increases the context size
  - Hard to answer because it uses convoluted language(*gandhi deeply influenced count wrote war peace*)
  - Resembles more keyword-based search queries than grammatical questions(importance of learning who , when type questions)

Improvements in BiDAF model :

- Embeddings from Language modelling rather than pretrained
- Using self - attention : incorporating multiple hops of attention to allow deeper interaction between context and query

Reinforcement Learning

- Model Pipeline
- Why RL?
- REINFORCE
- Model details
- Results
- Discussion
- Error Analysis

RL: Why?

- SearchQA has multiple passages/snippets for every query
- Correct answer can also be in multiple passages
- Gold passage not given: No Supervised Learning Setting possible
- Using Reinforcement Learning/ reward based objective for passage selection
- **Task:** Select one of the passages from which answer is to be selected
- **Reward:** High reward if the selected passage give high confidence score to the gold answer

RL: REINFORCE algorithm

- Objective Function: Maximize the expected reward $J(\theta)$

$$J(\theta) = \sum_{p_k \in \text{passages}} p_\theta(p = p_k | \text{query, passages}) R_\theta(p_k)$$

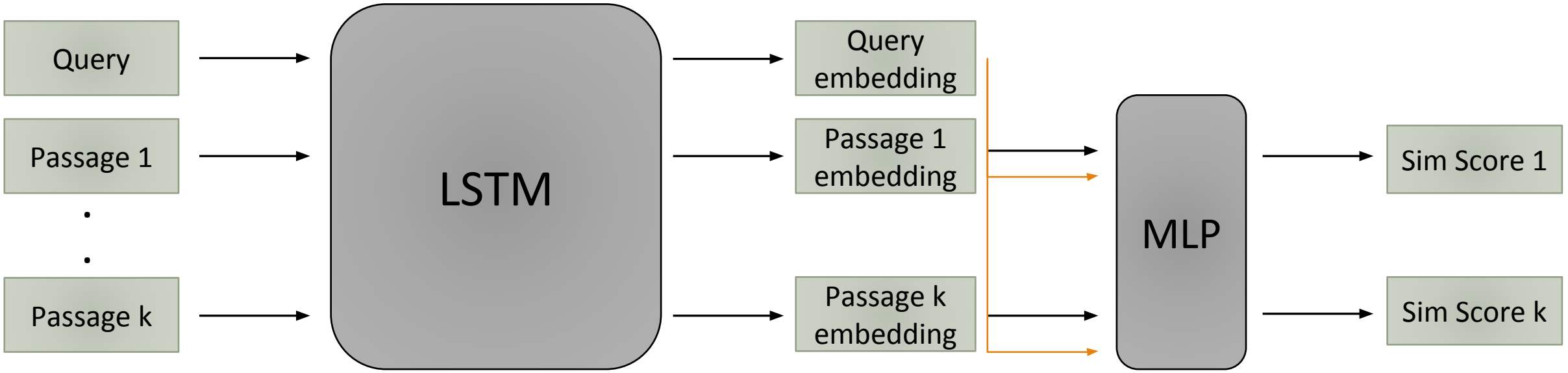
- $R_\theta(p_k) = p_1(y = y_{start}) * p_2(y = y_{end})$ or 0 if incorrect passage selected

  where $R_\theta(p_k)$ is the reward output from BiDAF model

- Gradient of the objective function can be approximated with a sample:

$$\nabla J(\theta) \approx \nabla \log p_\theta(y|\hat{p}, \text{query}) + \log p_\theta(y|\hat{p}, \text{query}) \cdot \nabla \log p_\theta(\hat{p}|\text{query, passages})$$

RL: Model Details



RL: Model Details - Sampling

- For each example in the batch, one passage is sampled based on the similarity scores
- Similarity scores and Sampling: Independent of the order of passages (and the number of passages)
- Training:
  - Supervised setting (where simple heuristic is used to select gold passage)
  - Reinforcement Learning
- Supervised setting: Pick the first passage with gold answer as the gold label and use negative log likelihood to train the RL model
- RL setting: Use the sampled passage and pass it through BiDAF to receive reward and train it using REINFORCE

RL: Training Details

- Baseline
  - To reduce variance during training and making it stable
  - Heuristic: Running average of reward kept as baseline score
  - Why: Only positive rewards in our case
  - Need to have positive and negative rewards for model to train
  - Actions (passages sampled) getting higher reward (than baseline) -> increase probability of those actions (passages)
  - Modified Reward: r - b (Actions that perform better on an average end up with positive reward)
- Training: For each batch example in epoch 'e': choose between SL/RL based on γ^e where γ = 0.8

RL: Results

Question	Gold Passage	Predicted Passage (Confidence Score)	Error Type
oil cartel controls 40 world production	oil producing exporting cartels act 2007 even though **opec** controls ``only'' 40 world 's production , influence prices substantial first , opec countries extensive reserves	• … 10 opec nations pump 40 world 's oil supplies … (0.2) • **opec** controls ``only'' 40 world 's production … (0.13) • … opec , cartel controls 40 world 's oil production …(0.25)	Too many passages with the gold answer
name texas city spanish yellow	**amarillo** , texas wikipedia amarillo 14th populous city state texas , united states also city also known yellow rose texas city takes name spanish word yellow , recently rotor city	• …texas city 's spanish yellow answers 8 letters texas city 's spanish … (0.3) • **amarillo**, … yellow rose texas city … (0.25) • amarillo texas cities traveltexas amarillo , means yellow spanish … (0.25)	Passage does not contain the correct answer but exact query
18 became queen 1837 , reigned 63 years	queen **victoria** biography undiscovered scotland last monarch house hanover , ruled 63 years 7 18th birthday , william iv died heart failure victoria became queen idea marriage two , course longer say 1837	• queen elizabeth ii make 63 years 217 days throne (0.1) • queen **victoria** … ruled 63 years 7 18th… (0.14) • queen victoria reigned 63 years , seven months , two days (0.23)	Wording + Noisy snippet

RL: Discussion

- **Reward:** Reward increases but not by a huge amount (high variance also)
- Around 40-50% queries have 15+ snippets containing the gold answer
  - Choosing gold passage heuristically makes training difficult
  - Alternative: Increasing the probability of all snippets containing the gold answer: multi-label approach
- Snippets very noisy: Hard to get the similarity scores correct, eg:
  - 7 surprising facts queen victoria history extra aug 26 , 2016 1 18 became queen went sitting room 6am 20 june 1837 , young princess woken bed course 63 year long reign , victoria came
  - iris 2001 imdb biography true story lifelong romance novelist iris murdoch husband husband john bayley , student days battle alzheimer 's disease portraying author poll image house , iris murdoch spent final years life , still south oxford
  - johnson 's universal cyclopaedia scientific popular treasury google books result

Conclusion

- Implemented an end-to-end QA pipeline with reward based objective function
- Baseline QA model for span selection given query and passage
- RL model for sampling a snippet (from multiple snippets) based on similarity between query and passage
- Few ideas to try:
  - Sample multiple passages (and see if training improves) – but can harm the performance also
  - Query reformulation: Reformulate the query so that the model can return correct answers accurately and confidently (https://arxiv.org/pdf/1705.07830.pdf)
- In conclusion: RL may not work so well for SearchQA in the current setting without any additional pre-processing or model adaptations

THANK YOU