High Performance SQuAD & Transfer Learning

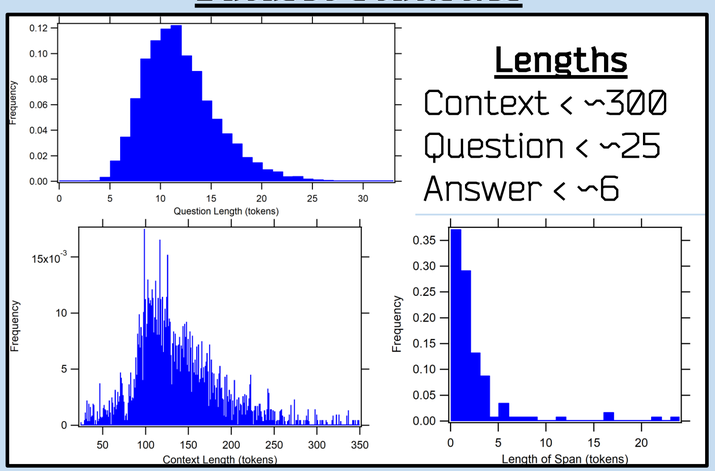
Jeff Chen and Alexandre Gauthier

Background

- Want to answer questions with solutions taken from a short passage
- SQuAD dataset: >100,000 examples from Wikipedia [1]
- State of the art: F1=0.89 [2]
- Extrapolate trained model to other types of data sources

Dataset & Baseline Model

Dataset Statistics



Lengths

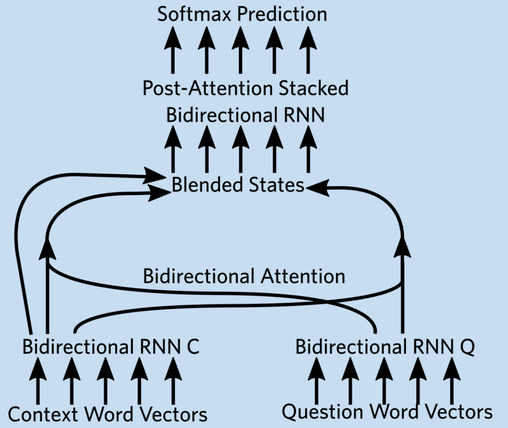
Context < ~300
Question < ~25
Answer < ~6

Baseline Model

- Bidirectional RNNs on context, question word vectors
- Attention from context to question
- Predict start, end position independently using separate fully connected layers / Softmax

jc1 / agau @stanford.edu

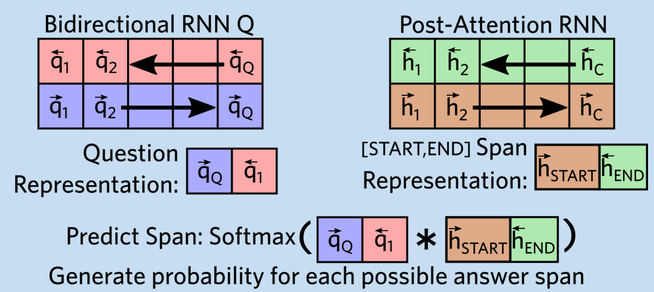
Main Model Modifications



Bidirectional Attention Flow – Attention flows both ways between context and question hidden states [3]

Post-Attention RNN – Add a new stacked bidirectional LSTM after the attention layer, with the blended reps as inputs

Span Representation – Prevent model from choosing impossible span



References

[1] P. Rajpurkar *et al.*, CoRR, abs/1606.05250 (2016)
[2] M. Hu *et al.*, arXiv:1705.02798v3 (2017)
[3] M. Seo *et al.*, arXiv:1611.01603v5 (2017)
[4] Y. Yu *et al.*, arXiv:1610.09996v2 (2016)
[5] ft.com/lex
[6] Corporate report & general news datasets compiled by K. Sill

Results

Model	Dev F1 Score
Baseline	0.403
Spans "Full"	0.521
Spans "Simplified"	0.442
BiDAF	0.484
BiDAF + Dropout	0.486
BiLSTM + BiDAF + D.O.	0.673
BiDAF + BiLSTM + Spans "Simplified"	**0.686 (DEV)** **0.743 (TEST)**

Translation to Other Data Sources

- 17/25=0.68 correct on financial news [5]
- 22/30=0.73 correct on corporate reports
- 27/30=0.90 correct on general news [6]

Error Analysis

Examined model performance by hand

Common correct cases

- "How many…," "How long…" questions
- Structure of question is similar to text surrounding answer in the context

Common wrong cases

- Picking items from a list based on reasoning, e.g. "Which other movie…" or "Besides X…"
- Predicted answers are too long
- Answer requires prior knowledge of format, e.g. "Which director…" wants a person's name