MACHINE COMPREHENSION USING MATCH-LSTM AND ANSWER POINTER

Nikitha Pasya (B00830710)

Introduction

- An end-to-end neural architecture for the question answering task.
- The architecture is based on match-LSTM and Pointer Net, a sequence-to-sequence model to constrain the output tokens to be from the input sequences.
- Given a piece of text, which is referred to as a passage, and a question related to the passage. The goal is to identify a subsequence from the passage as the answer to the question.

SQuAD

- SQuAD provides a challenging testbed for evaluating machine comprehension algorithms.
- In SQuAD the answers do not come from a small set of candidate answers and they have variable lengths.
- The questions and answers in SQuAD were created by humans through crowdsourcing, which makes the dataset more realistic.

In 1870, Tesla moved to Karlovac, **to attend school at the Higher Real Gymnasium**, where he was profoundly influenced by a math teacher **Martin Sekulic´**. The classes were held in **German**, as it was a school within the Austro-Hungarian Military Frontier. Tesla was able to perform integral calculus in his head, which prompted his teachers to believe that he was cheating. He finished a four-year term in three years, graduating in 1873.

---

1. In what language were the classes given?	German
2. Who was Tesla’s main influence in Karlovac?	Martin Sekulic´
3. Why did Tesla go to Karlovac?	attend school at the Higher Real Gymnasium

MATCH-LSTM

- To predict whether the premise(question) entails the hypothesis(passage), the match-LSTM model goes through the tokens of the hypothesis sequentially.
- At each position of the hypothesis, attention mechanism is used to obtain a weighted vector representation of the premise.
- Weighted premise is then to be combined with a vector representation of the current token of the hypothesis and fed into an LSTM, which we call the match-LSTM.
- Match-LSTM sequentially aggregates the matching of the attention-weighted premise to each token of the hypothesis and uses the aggregated matching result to make a final prediction.

Ptr-Net

- Ptr-Net uses attention mechanism as a pointer to select a position from the input sequence as an output symbol.
- To construct answers using tokens from the input text.
- Two ways of using Pointer Net for QA.
    - Sequential model
    - Boundary model

Overview of Neural network model

- The passage is represented by matrix P ∈ R^{d*P} and the question is represented by matrix Q ∈ R^{d*Q}
- LSTM Preprocessing Layer

  standard one-directional LSTM to preprocess the passage and the question separately

  $H^p = \overrightarrow{LSTM}(P)$, $H^q = \overrightarrow{LSTM}(Q)$
- Match-LSTM Layer
  - tries to match the passage against the question.
  - At position i of the passage, it first uses the standard word-by-word attention mechanism to obtain attention weight vector $\overrightarrow{\alpha_i} \in R^Q$ as follows:

    $\overrightarrow{G_i}$ = tanh($W^q H^q + (W^p h_i^p + W^r \overrightarrow{h_{i-1}^r} + b^p) \otimes e_Q$)

    $\overrightarrow{\alpha_i}$ = softmax($w^T \overrightarrow{G_i}$ + b $\otimes e_Q$)

use the attention weight vector $\overrightarrow{\alpha_i}$ i to obtain a weighted version of the question and combine it with the current token of the passage to form a vector $\overrightarrow{z_i}$.

- $\overrightarrow{z_i}$ is fed into a standard one-directional LSTM to form match-LSTM:

$$\overrightarrow{h_i^r} = \overrightarrow{LSTM}(\overrightarrow{z_i}, \overrightarrow{h_{i-1}^r})$$

where $\overrightarrow{h_i^r} \in R^{(l)}$

- further build a similar match-LSTM in the reverse direction

$$\overleftarrow{G_i} = \tanh(W^q H^q + (W^p h_i^p + W^r \overleftarrow{h_{i+1}^r} + b^p) \otimes e_Q)$$

$$\overleftarrow{\alpha_i} = \text{softmax}(w^T \overleftarrow{G_i} + \text{b} \otimes e_Q)$$

- Let $\overrightarrow{H^r} \in R^{(l * P)}$ represent the hidden states $[\overrightarrow{h_1^r}, \overrightarrow{h_2^r}, \overrightarrow{h_3^r}, \ldots\ldots, \overrightarrow{h_p^r}]$ and $\overleftarrow{H^r} \in R^{(l * P)}$ represents $[\overleftarrow{h_1^r}, \overleftarrow{h_2^r}, \overleftarrow{h_3^r}, \ldots\ldots, \overleftarrow{h_p^r}]$

$H^r \in R^{(2l * P)}$ is the concatenation of two $\begin{bmatrix} \overrightarrow{H^r} \\ \overleftarrow{H^r} \end{bmatrix}$

- **Answer Pointer Layer**
  - uses Ptr-Net to select a set of tokens from the passage as the answer.
  - This layer uses the sequence H^r as input.

  The Sequence Model: The answer is represented by a sequence of integers a = (a1, a2, . . .) .
  - Allows each a_k to take up an integer value between 1 and P + 1, Once a_k = P + 1, the generation of the answer stops.
  - To generate the answer token a_k, the attention mechanism is used again to obtain an attention weight vector $\beta_k \in R^{(P+1)}$, where $\beta_{k,j}$(1 ≤ j ≤ P + 1) is the probability of selecting the j th token from the passage as the k th token in the answer, β_k is modeled as follows:

$$F_k = \tanh(V\,\widetilde{H}^r + (W^a\, h^a_{k-1} + b^a) \otimes e_{(P+1)}),$$

$$\beta_k = \mathrm{softmax}(V^T F_k + c \otimes e_{(P+1)})$$

Where $\widetilde{H}^r$=[H^r;0]

  - We can then model the probability of generating the answer sequence as

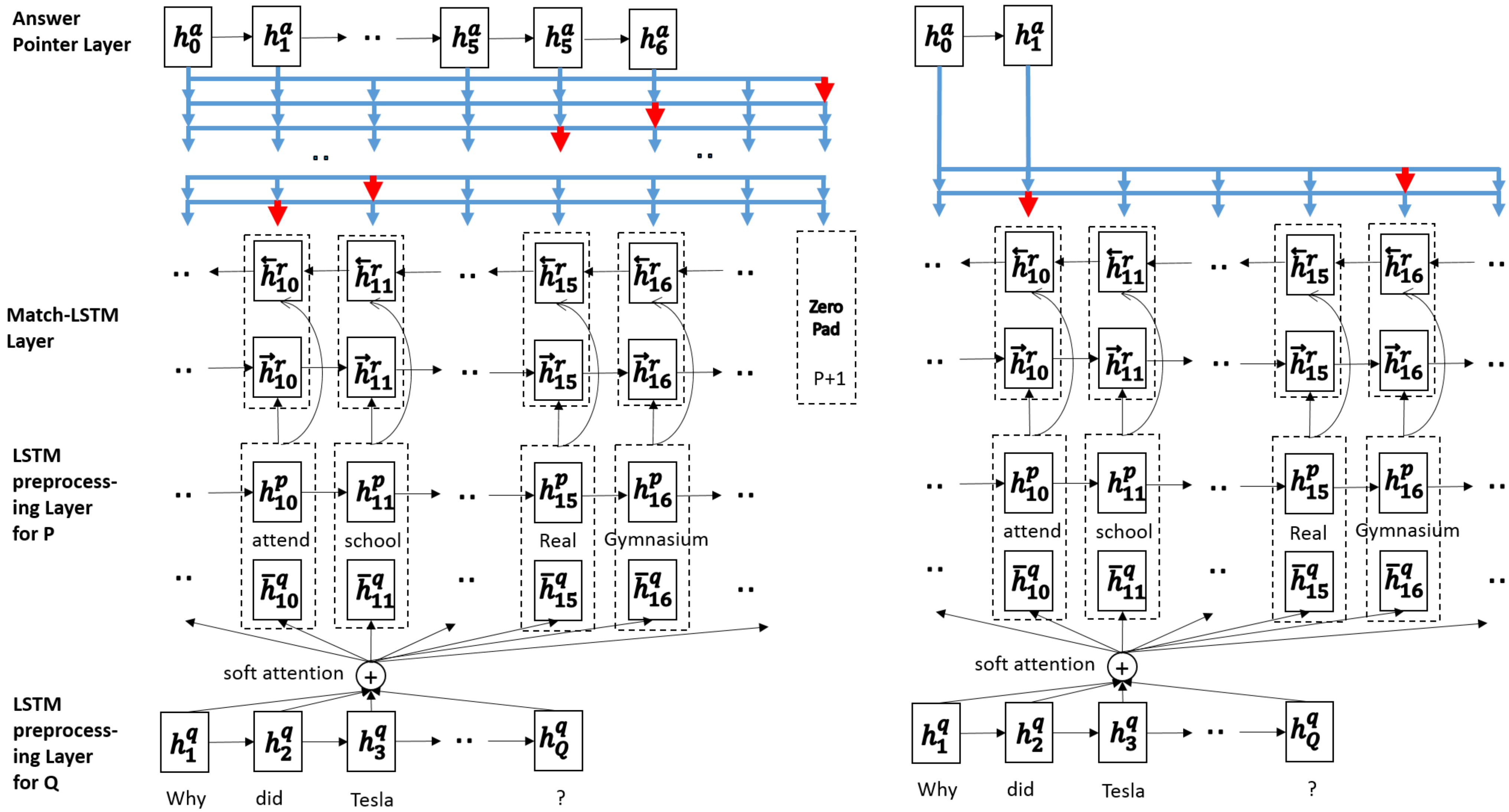
$$p(a|H^r) = \prod_k p(a_k|a1, a2, \ldots, a_{k-1}, H^r)$$

$$p(a_k = j|a1, a2, \ldots, a_{k-1}, H^r) = \beta_{k,j}$$

- **The Boundary Model:** Predicts two indices a_s and a_e.
  - the probability of generating an answer is simply modeled as

$$p(a|\ H^r\ ) = p(a_s\ |\ H^r)p(a_e|\ a_s\ ,H^r)$$

  - the boundary model is further extended by incorporating a search mechanism. During prediction, the length of the span is limited and globally search the span with the highest probability computed by $p(a_s) \times p(a_e)$.
  - As the boundary has a sequence of fixed number of values, bi-directional Ans-Ptr can be simply combined to fine-tune the correct span.



(a) Sequence Model

(b) Boundary Model

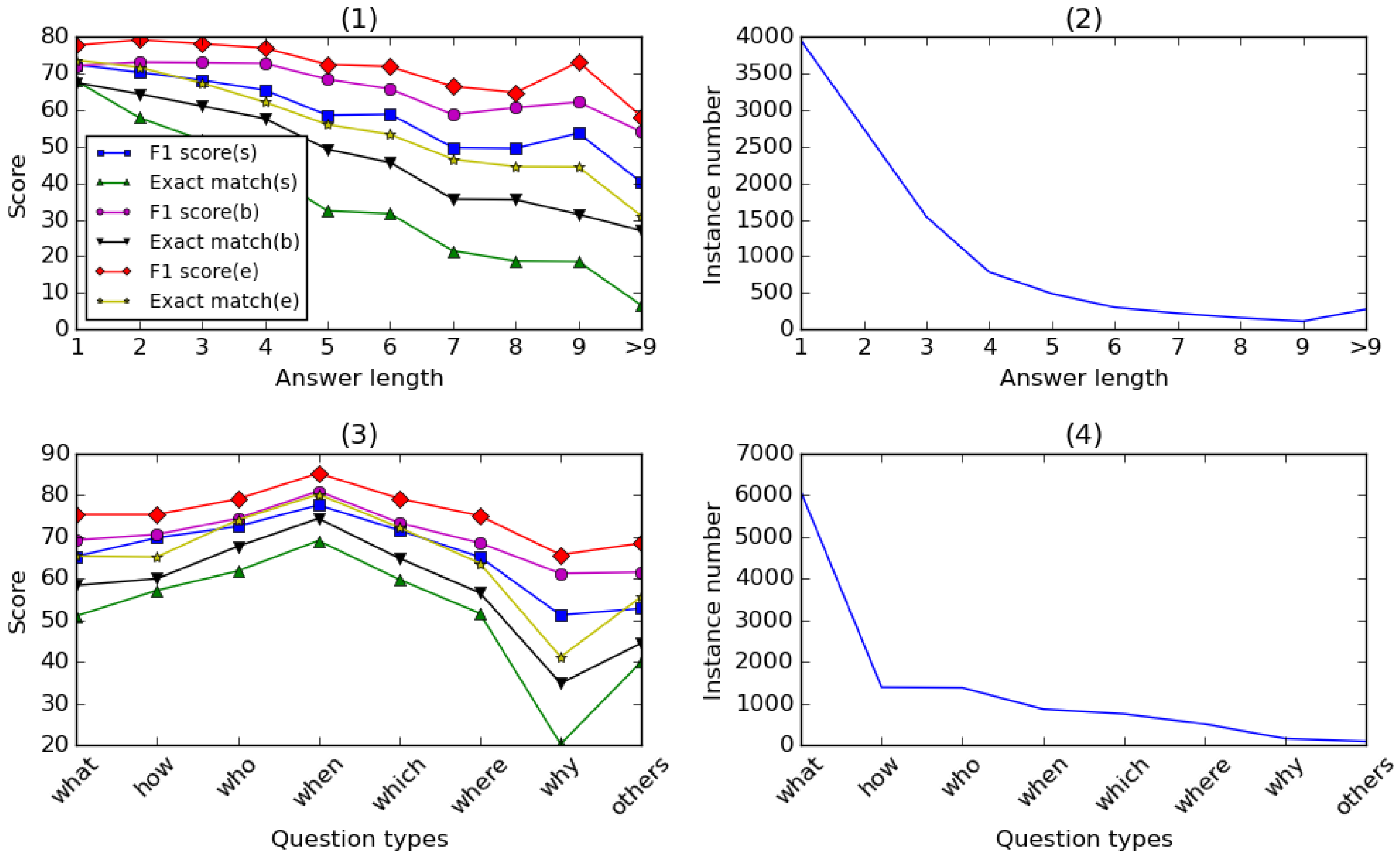
Experiments

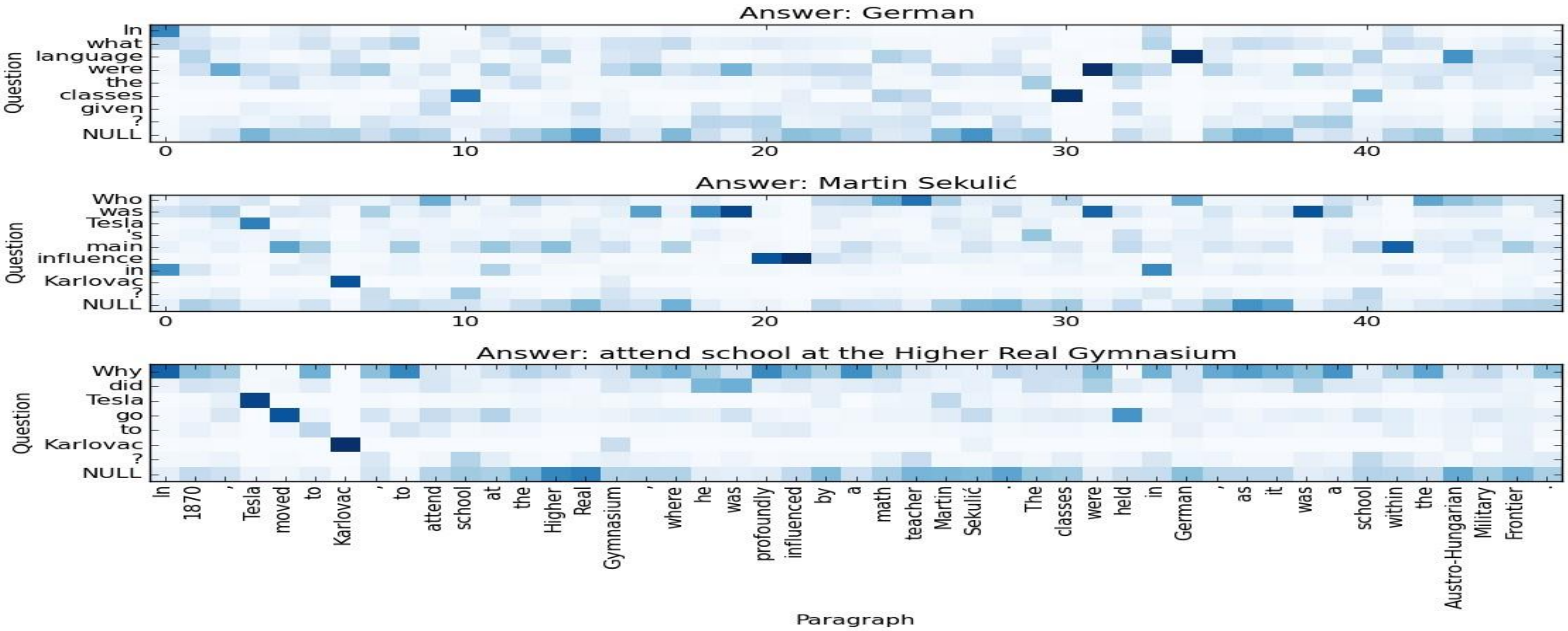
- The dimensionality l of the hidden layers is set to be 150 or 300.
- We use ADAMAX with the coefficients β1 = 0.9 and β2 = 0.999 to optimize the model. Each update is computed through a minibatch of 30 instances.
- The performance is measured by two metrics: percentage of exact match with the ground truth answers, and word-level F1 score when comparing the tokens in the predicted answers with the tokens in the ground truth answers.
- F1 scores with the best matching answers are used to compute the average F1 score.

Results

	l	$\|\theta\|$	Exact Match Dev	Exact Match Test	F1 Dev	F1 Test
Random Guess	-	0	1.1	1.3	4.1	4.3
Logistic Regression	-	-	40.0	40.4	51.0	51.0
DCR	-	-	62.5	62.5	71.2	71.0
Match-LSTM with Ans-Ptr (Sequence)	150	882K	54.4	-	68.2	-
Match-LSTM with Ans-Ptr (Boundary)	150	882K	61.1	-	71.2	-
Match-LSTM with Ans-Ptr (Boundary+Search)	150	882K	63.0	-	72.7	-
Match-LSTM with Ans-Ptr (Boundary+Search)	300	3.2M	63.1	-	72.7	-
Match-LSTM with Ans-Ptr (Boundary+Search+b)	150	1.1M	63.4	-	73.0	-
Match-LSTM with Bi-Ans-Ptr (Boundary+Search+b)	150	1.4M	**64.1**	**64.7**	**73.9**	**73.7**
Match-LSTM with Ans-Ptr (Boundary+Search+en)	150	882K	**67.6**	**67.9**	**76.8**	**77.0**

Further Analysis




Answer: German
Answer: Martin Sekulić
Answer: attend school at the Higher Real Gymnasium
Question
In what language were the classes given ? NULL
Who was Tesla 's main influence in Karlovac ? NULL
Why did Tesla go to Karlovac ? NULL
0 10 20 30 40
In 1870 , Tesla moved to Karlovac , to attend school at the Higher Real Gymnasium , where he was profoundly influenced by a math teacher Martin Sekulić . The classes were held in German , as it was a school within the Austro-Hungarian Military Frontier .
Paragraph

Conclusion

- Developed two models for the machine comprehension problem defined in the Stanford Question Answering (SQuAD) dataset, both making use of match-LSTM and Pointer Network.
- Experiments on the SQuAD dataset showed that the boundary model, could achieve an exact match score of 67.6% and an F1 score of 77% on the test dataset.

References

- S.Wang and J.Jiang.Machine Comprehension using Match-LSTM and Answer pointer. arXiv preprint arXiv:1608.07905, 2017.
- Oriol vinyals, Meire Fortunato ,Navdeep Jaitly.Pointer Networks.arXiv preprint arxiv.org:1506.03134,2017.