Question Answering Kaggle Competition

Misha Yaschenko
Yana Shishkina

Question Answering Task

The goal is to predict **short** and **long** answer responses to real questions about Wikipedia articles

Long answer

Sequence classification task
Classify each given paragraph
if it's the right answer.

Yes/No answer

Sequence classification task
Classify each long answer & short question if it has yes/no answer or it's not applicable

Short answer

Token classification task
Classify each token of the right long answer if it belongs to the short answer

Question Answering Task

BERT

Bert-base-cased
Pretrained

F1 Micro

Chosen by Kaggle

PyTorch

PyTorch

Long Answer

Idea

- clean texts from tags
- we clean the texts from stop words (later refused)
- divide the texts from wikipedia into paragraphs, and match each paragraph with a target
- predict the most probable positive candidate from each document

Problems

- unbalanced sampling: there is only one paragraph for a large text containing the answer

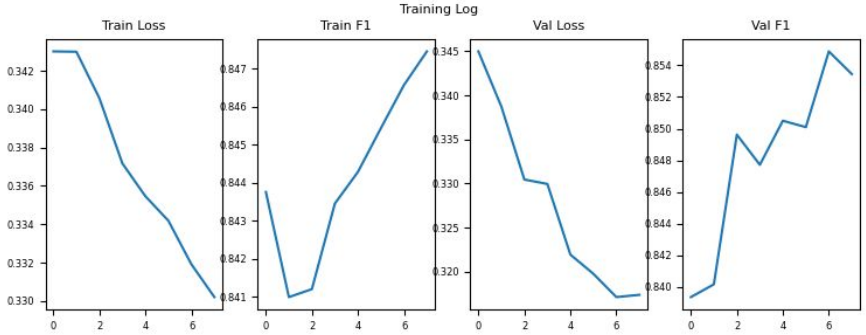
What else could be taken into account

- the possible usefulness of tags
- calibration of probabilities
- try other models from the same family of models (for example bert-large)

(question, paragraph) -> is_long_answer

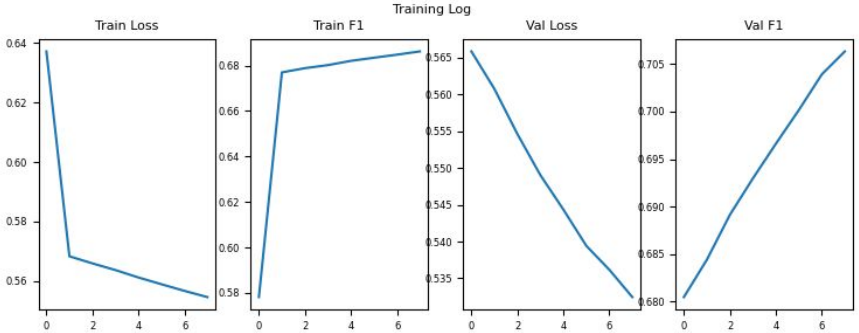
Long Answer

2 trainable BERT layers (2nd epoch log)



On test (real formatting):
F score: 57.57

fully frozen BERT (2nd epoch log)

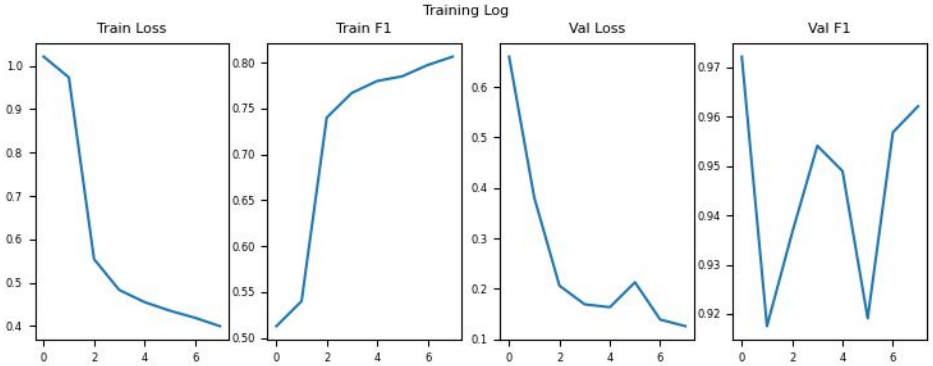


Yes/No Answer

3-class classification

97.5% - NONE!

Balanced to:
NONE: 4000
NO: 1439
YES: 2359



<u>Our best</u>
F score: 95.74

<u>Majority baseline</u>
F score: 97.67

:(

Short Answer

Idea

- Split long answers in parts (a bit smaller than max_len)
- Classify each token if it belongs to answer (2 classes)
- Map Bert tokenizer tokens to regular tokens

Problems

- Short answer can span multiple parts of long answer
- Can predict multiple answers in one long answer
- Unknown tokens can push the answer out of max_len bounds

(question, long_answer) -> short_ans_indices

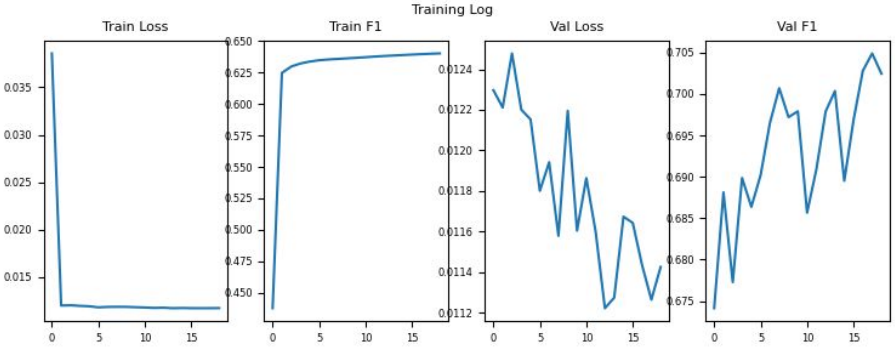
Short Answer

Example

Long answer: *The quick brown fox jumps over the lazy dog*
Question: *What are the characteristics of the dog?*
Answer: *7:8*
Max len: 5
Parts: [*The quick brown fox jumps, over the lazy dog*]
Model output: [-1:-1, 2:3]
Aligned output: 2+5 : 3 + 5 → 7:8

Short Answer

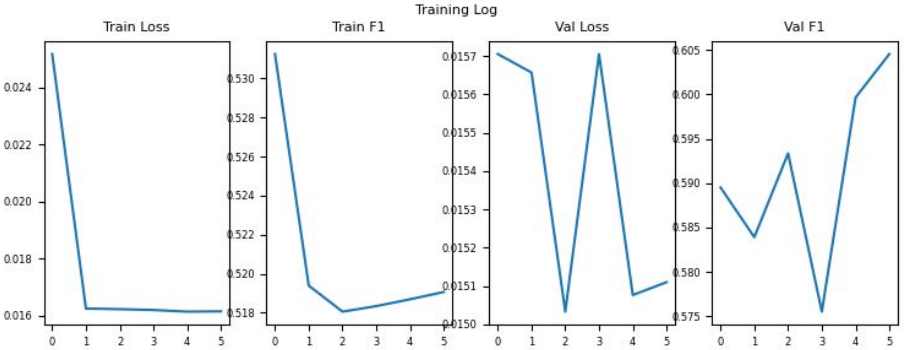
fully trainable BERT (2nd epoch log)



On test (real formatting):
F score: 44.50

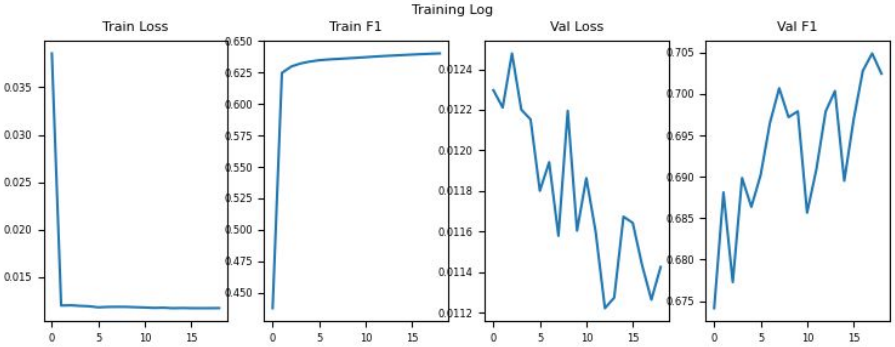
On test (our formatting):
F score: 68.00

2 trainable BERT layers (4th epoch log)



Short Answer

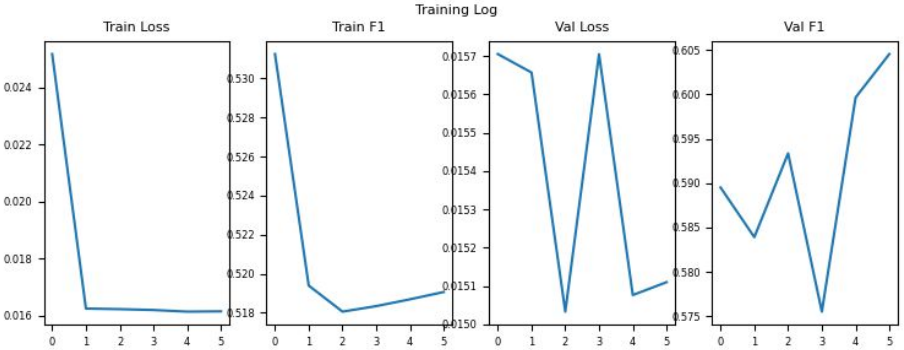
fully trainable BERT (2nd epoch log)



On test (real formatting):
F score: 44.50

On test (our formatting):
F score: 68.00

2 trainable BERT layers (4th epoch log)



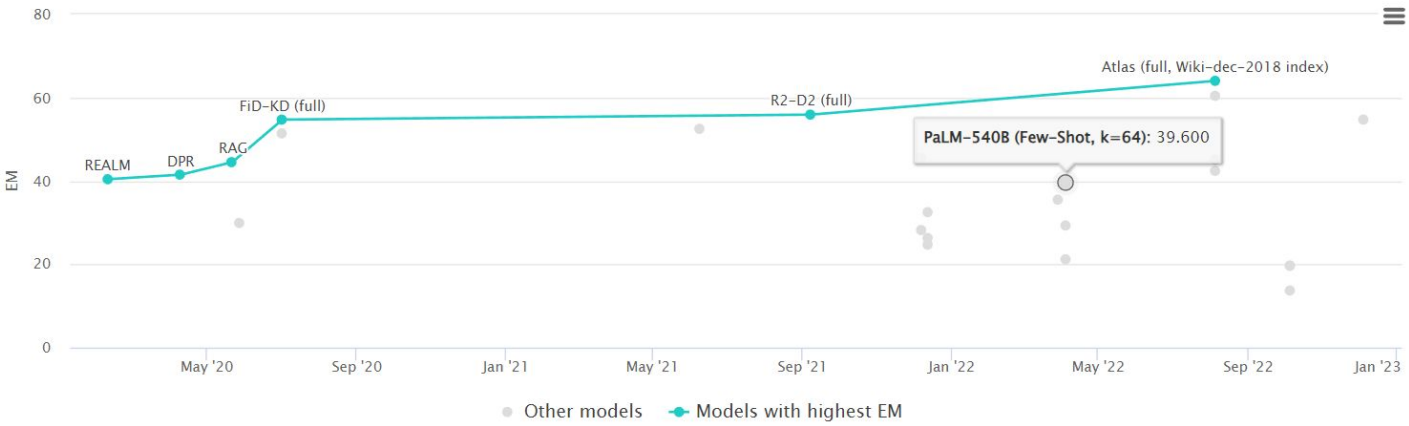
We are not alone!

PaLM – 540B BERT – 110M

Question Answering on Natural Questions

Leaderboard Dataset

View EM by Date



Kaggle Score

Challenges with Kaggle:

- very large test sample and Kaggle memory limit
- gluing models
- bert-base is downloaded from the Internet ;(
- GPU limitation
- very long submission

Public

939	Prince MAkwana	0.21993	2	3y
940	Marek Swieton	0.21324	17	3y
	TF2.0 Bert Baseline	0.21286		

Private

921	▼ 19	Bill Miao	0.24017	1	3y
922	▼ 39	senkin13	0.23807	2	3y
		TF2.0 Bert Baseline	0.23723		

Submission and Description	Private Score	Public Score
qa notebook x 2 - Version 1 Succeeded (after deadline) · 11h ago · janas version	**0.2472**	**0.22184**

Input

- tensorflow2-question-answering
- bert-base-cased
- long-answer-model
- transformers
- short-model-qa

REPO

https://github.com/yashkens/MLDM_QA_project