Group 28

Enhancing Tiny Transformers with NER & POS tagging for QA

Bogdan Bîndilă
m.bindila@student.utwente.nl

Cristina Racoviță
c.racovita@student.utwente.nl

Introduction

Question Answering has seen significant advances, primarily driven by **large transformer models**. Our study investigates the potential for enhancing smaller transformer models by incorporating external features - **NE and POS tags**.

Research Questions

Can the performance of compact models for **extractive QA with factoid answers** be improved?

What are the most **effective approaches** that do **not modify the model architecture?**

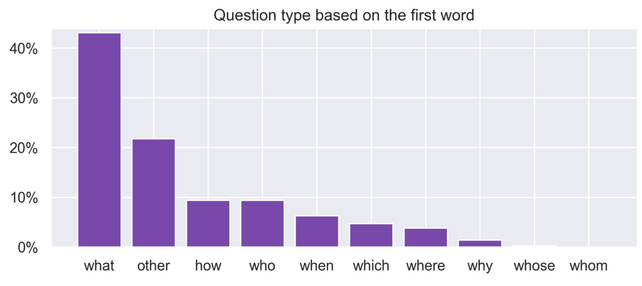
Data

100,000+ questions-answer pairs created by crowdworkers based on Wikipedia articles [1]

SQuAD

Question: What organization did Tesla serve as vice president of?

Context: Tesla served as a vice president of the American Institute of Electrical Engineers.



Added features

flair Based on **Flair embeddings** & LSTM-CRF [2]

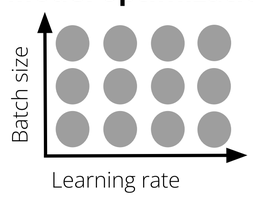
	4-Class NER	18-Class NER	POS
Model	ner-fast	ner-ontonotes-fast	pos
F1 score	92.75%	89.27%	98.19%
Training Dataset	Conll-03	Ontonotes	Ontonotes

Methods

Data pre-processing

- split question-context pairs into fixed-size chunks
- store answer start and end indices

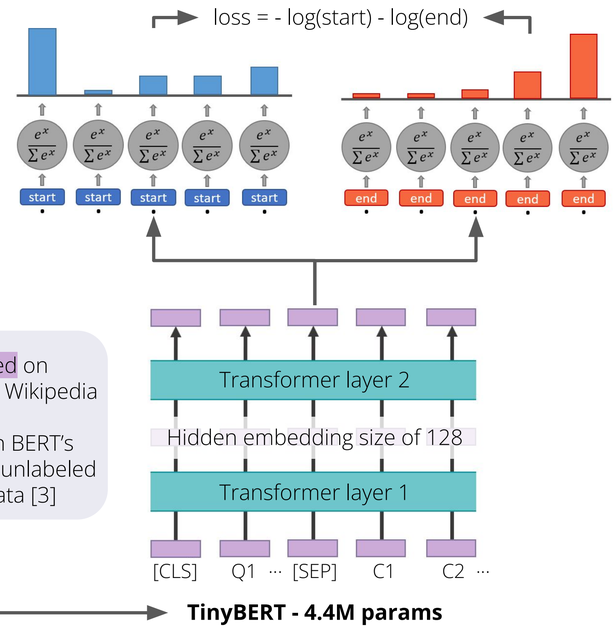
Model optimization



Data post-processing

- get the start & end index that maximize log-likelihoods sum

Fine-tuning



Pre-trained on BookCorpus & Wikipedia + Distilled with BERT's knowledge on unlabeled transfer data [3]

TinyBERT - 4.4M params

Concatenation experiments

POS

word1	pos1	word2	pos2	word3	pos3
word1	word2	word3	pos1	pos2	pos3

NER

word1	ne1	word2	ne2	word3	ne3
START ne1	START ne2	word1	word2	END ne2	END ne1

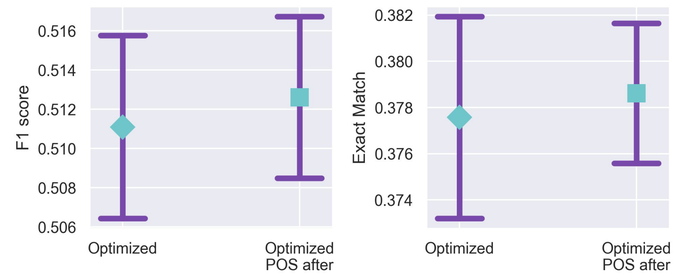
4/18 classes using one or multiple tags [4]

Results & Next Steps

Model metrics with 95% confidence intervals

Fine-tuned baseline EM: 0.3078 ± 0.0084
F1: 0.4305 ± 0.0092

Performance



After conducting all experiments

Analysis

What happens to the probability of the predicted answers?

How does performance change for different question types?

Hypothesis:

Prob. of correct answers ↑

Prob. of incorrect answers ↓

References

[1] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100,000+ questions for machine comprehension of text," arXiv preprint arXiv:1606.05250, 2016.

[2] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, "FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP," in Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations), pp. 54–59, 2019.

[3] I. Turc, M.-W. Chang, K. Lee, and K. Toutanova, "Well-read students learn better: On the importance of pre-training compact models," arXiv preprint arXiv:1908.08962, 2019.

[4] D. Moll´a, M. Van Zaanen, and D. Smith, "Named entity recognition for question answering," in Proceedings of the Australasian language technology workshop 2006, pp. 51–58, 2006.

UNIVERSITY OF TWENTE.