Develop New Transformer Architecture For Question and Answering(QandA)

Nirbhay P. Tandon

June, 2021

Contents

List of Figures **3**

List of Tables **4**

List of Abbreviations **5**

1 Introduction **6**

1.1 Background Of The Study 6
1.2 Aims And Objectives 6
1.3 Scope Of The Study 6
1.4 Significance Of The Study 6
1.5 Structure Of The Study 6

2 Literature Review **7**

2.1 Question Answering Using Neural Nets 7
2.2 Question Answering Using LSTMs 7
2.3 Question Answering Using Transformers 8
2.4 Comparison Of Techniques 8
2.5 Summary 8

3 Research Methodology **9**

3.1 Data Selection 9
3.2 Data Pre-processing And Transformation 9
3.3 Existing Models And Benchmarks 9

4 Architecture Creation **10**

4.1 Drawbacks Of Current Architectures 10
4.2 Proposed Architecture Improvements 10

4.3 Architecture Refinement 10

Appendices **11**

Bibliography **12**

List of Figures

List of Tables

List of Abbreviations

BPPTs Back-Propagation Through Time

LSTMs Long Short-Term Memory Architectures

MCTest Machine Comprehension of Text

NLP Natural Language Processing

NLU Natural Language Understanding

QASENT A challenge dataset for open-domain question answering

RACE ReAding Comprehension Dataset From Examinations

RNNs Recurrent Neural Networks

Chapter 1
Introduction

1.1 Background Of The Study

1.2 Aims And Objectives

1.3 Scope Of The Study

1.4 Significance Of The Study

1.5 Structure Of The Study

Chapter 2

Literature Review

2.1 Question Answering Using Neural Nets

2.2 Question Answering Using LSTMs

To be able to accurately assess the aims of this thesis that have been highlighted in 1.2, we must take a deeper look at the work done in the field of Long Short-Term Memory Architecture based models and their importance in NLP.

In this part of the literature review, our purpose is to highlight how LSTMs came to be & their particular usage in the field of Question Answering.

1. In 1997 Schmidhuber and Hochreiter (1997) introduced to the world a gradient descent based model in the form of Long Short-Term Memory. They set out to solve the problem of the vanishing gradient which was often seen in "Back-Propagation Through Time"(BPPTs) based neural networks. Extensive studies done on this, some by Hochreiter himself, showed that the problem of vanishing gradients is a real one. It was also seen that in case of BPTTs, the error back-flow mechanisms would either blow up or also suffer from vanishing gradients leading to either oscillating weights or learning to bridge long time gaps would not work. To remediate this, the authors introduced the concept of a *constant* error flow through the internal states of special units.

2. Hochreiter et al. (2001)

3. Schmidhuber (2015)
4. Tan et al. (2015)
5. Wang and Jiang (2016)
6. Gennaro et al. (2020)
7. Weissenborn et al. (2017)

2.3 Question Answering Using Transformers

1. Vaswani et al. (2017)
2. Devlin et al. (2018)
3. Lan et al. (2019)
4. Liu et al. (2019)
5. Sanh et al. (2019)
6. Rajpurkar
7.
8.

2.4 Comparison Of Techniques

2.5 Summary

(Corsair, 2021)

Chapter 3

Research Methodology

DONT DO ANY ANALYSIS HERE

3.1 Data Selection

3.2 Data Pre-processing And Transformation

3.3 Existing Models And Benchmarks

Chapter 4
Architecture Creation

4.1 Drawbacks Of Current Architectures

4.2 Proposed Architecture Improvements

4.3 Architecture Refinement

Appendices

Bibliography

Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*, 2019.

M. Tan, B. Xiang, and B. Zhou. Lstm-based deep learning models for non-factoid answer selection. *CoRR*, abs/1511.04108, 2015. URL `http://arxiv.org/abs/1511.04108`.

S. Wang and J. Jiang. Machine comprehension using match-lstm and answer pointer. *CoRR*, abs/1608.07905, 2016. URL `http://arxiv.org/abs/1608.07905`.

V. Sanh, L. Debut, J. Chaumond, and T. Wolf. Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter. *CoRR*, abs/1910.01108, 2019. URL `http://arxiv.org/abs/1910.01108`.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Corsair. Vengeance® lpx 8gb (1 x 8gb) ddr4 dram 2400mhz c14 memory kit - black, 2021. URL `https://www.corsair.com/uk/en/Categories/Products/Memory/VENGEANCE-LPX/p/CMK8GX4M1A2400C14`. Accessed: 2021-04-16.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

G. D. Gennaro, A. Buonanno, A. D. Girolamo, A. Ospedale, and F. A. N. Palmieri. Intent classification in question-answering using LSTM architectures. *CoRR*, abs/2001.09330, 2020. URL `https://arxiv.org/abs/2001.09330`.

P. Rajpurkar. Squad2.0. URL `https://rajpurkar.github.io/SQuAD-explorer/`.

S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.

J. Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

D. Weissenborn, G. Wiese, and L. Seiffe. Fastqa: A simple and efficient neural architecture for question answering. *CoRR*, abs/1703.04816, 2017. URL `http://arxiv.org/abs/1703.04816`.

J. Schmidhuber and S. Hochreiter. Long short-term memory. *Neural Comput*, 9(8):1735–1780, 1997.