Exposé for Master thesis

Zero-shot Cross-Lingual Transfer for Text Simplification with Adapters ł

Julian Pollmann, B.Sc.
Applied Cognitive and Media Science
julian.pollmann@uni-due.de

Supervisors
Prof. Dr. Daniel Bodemer
Prof. Dr. Christin Seifert
Jan Trienes, M.Sc.

UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

v0.1

1 INTRODUCTION

Machine Learning (ML) approaches for Natural Language Processing tasks like Text Simplification (TS) often require a large amount of data during training. Since in some languages more training data is available Machine Learning models trained on this data can perform better than models trained on low resource languages. In this context one approach could be to take advantage of what a model from a high resource language learned and apply this to the domain of the low resource language, which is called Transfer Learning. Adapters for Transformer based architectures are a recent approach that promise to tackle this issue. The aim of this master thesis is to investigate how suitable adapters are to solve this cross-lingual transfer and to examine potential drawbacks to this approach.

1.1 Motivation

Recent developments in the field of Natural Language Processing (NLP), such as ChatGPT, have demonstrated the power of these tools. Machine Learning based approaches drive Tools for Chatbots, Machine Translation or Text Simplification, and they have in common that they try to support humans in their tasks. Text Simplification, for example, aims to make a Text simpler and better to understand while preserving its meaning. Simplified Texts can assist users with reading difficulties, language learners or people with other challenges. Models for Text Simplification are commonly created by fine-tuning a pretrained model with a large amount of textual data. But data for this kind of task is sparse and is even less available in low resource languages, like e.g. German. Manually creating this training data is costly and time-consuming and retraining in every language is very inefficient and slow.

1.2 Need for Research

One solution to tackle the data sparsity problem is to use transfer learning, where in the case of Text Simplification learnings from a high resource language can be transferred to a low resource language. Adapters for Transformer based architectures are a recent approach that promise to enable this transfer and encapsulate the learnings for easier model sharing. Additionally they achieve performance similar to completely fine-tuned models on most tasks. [8] However it is questionable whether the quality of Text Simplifications with adapters is on par with that of fully fine-tuned models. Also further research is necessary, if the transfer from a high resource language to a low resource language yields comparable results. Therefore this thesis will investigate following aspects of adapters.

1.3 Research Questions

Given the previous undiscovered aspects of adapters following research questions could be made:

How effective are adapters for cross-lingual transfer on text simplification tasks?

How effective are adapters for cross-lingual transfer on text simplification tasks?
What are the shortcomings of using adapters as a way of cross-lingual transfer?

2 RELATED WORK

2.1 Text Simplification

According to Alva-Manchego et. al "Text Simplification (TS) is the task of modifying the content and structure of a text in order to make it easier to read and understand, while retaining its main idea and approximating its original meaning". [1] The simplified version of a text can assist e.g. non-native speakers, children or non-expert readers in their reading ability. [1, 6]

Text Simplification can be considered as a Sequence-to-Sequence Task, meaning a given input sequence of text can be mapped to an output sequence of text, where the output sequence is a simplified version of the input. [3] Therefore, some authors argue that TS is a monolingual translation task, where a complex input is translated into a simpler version. [2] In addition to Machine Translation, Text Summarization is closely related to the Text Simplification task. Most recent approaches to solve these Natural Language Processing (NLP) tasks are using

Transformers, a deep-learning model based on an Encoder-Decoder architecture. To train these models large amounts of data is necessary, for which reason pre-trained transformer models are commonly fine-tuned to a specific task. Fine-tuning is done using large parallel corpora. In the case of fine-tuning a text simplification model a large set of complex-simple sentence or document pairs is used, depending on the granularity of the model to train. [7] During training the Text Simplification model learns what modifications need to be made to transform a complex input into a simple one. The transformations done by a trained sentence Text Simplification model "range from replacing complex words or phrases for simpler synonyms, to changing the syntactic structure of the sentence (e.g., splitting or reordering components)". [1] This means that TS models can reduce lexical and syntactic complexity and explain complex concepts. [10]

2.2 Adapters in the Transformer architecture

Fine-tuned models depend on very large amounts of data and their performance increases with their number of parameters. [8] This leads to models with billions of parameters, making it parameter inefficient. [4] Therefore every fine-tuned model fits only one NLP task, and it is expensive to save and distribute these models. This also makes composing different tasks more difficult. [8] Adapters avoid this issues by placing layers in between the layers of a deep-neural net in a pretrained model. The weights of the in between layers are updated during training phase, while the weights in the layers of the pretrained model stay unchanged. Accordingly sharing the parameter weights is easier, e.g. to use them for another task. [4] This makes Adapters suitable for transfer learning regard to different tasks or in cross-lingual contexts. Adapterhub is a framework for building and integrating Adapters for Transformer-based language models and is based on the Huggingface Transformers library. [5, 9]

2.3 Transfer Learning

As previously mentioned fine-tuned models are tailored towards a specific Task and/or language, e.g. Text Simplification.

3 METHODOLOGY

3.1 Datasets

* Newsela * Definition + Manual + Auto * Language, Granularity Statistics (Number of Articles, Number of Sentences/Instances, Readabiliy Levels)

3.2 Baseline Text Simplification models

Fine-tune Reference Text Simplification model(s) * Wofür soll die Baseline genutzt werden? * Huggingface Trainer Api * Encoder-Decoder architecture * Complex-Simple as Inputs

3.3 Adapters for Text Simplification

Fine-tune Reference Text Simplification model(s)

3.4 Evaluation

Metrics (Which metrics to use) * Simplification vs. Automatic Readabiliy Assessment * SARI * BLEU * FKGL * EASSE Library

4 PROJECT PLANNING

4.1 Milestones

REFERENCES

[1] Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2020. Data-Driven Sentence Simplification: Survey and Benchmark. *Computational Linguistics* 46, 1 (November 2020), 135–187. https://doi.org/10.1162/coli_a_00370

[2] Dennis Aumiller and Michael Gertz. 2022. Klexikon: A German Dataset for Joint Summarization and Simplification. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 2693–2701.

[3] Yoav Goldberg. 2017. *Neural Network Methods for Natural Language Processing*. Number 37 in Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, San Rafael, Calif. https://doi.org/10.2200/S00762ED1V01Y201703HLT037

[4] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-Efficient Transfer Learning for NLP. In *International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*. PMLR, 2790–2799.

[5] Huggingface. 2023. Huggingface - The AI community building the future. https://huggingface.co/. Accessed: 2023-01-22.

[6] Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. 2020. Neural CRF Model for Sentence Alignment in Text Simplification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 7943–7960. https://doi.org/10.18653/v1/2020.acl-main.709

[7] Daniel Jurafsky and James H. Martin. 2023. Speech and Language Processing (3rd Ed. Draft). https://web.stanford.edu/~jurafsky/slp3/. Accessed: 2023-01-22.

[8] Jonas Pfeiffer, Andreas Rücklé, Clifton Poth, Aishwarya Kamath, Ivan Vulić, Sebastian Ruder, Kyunghyun Cho, and Iryna Gurevych. 2020. AdapterHub: A Framework for Adapting Transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online, 46–54. https://doi.org/10.18653/v1/2020.emnlp-demos.7

[9] Jonas Pfeiffer, Andreas Rücklé, Clifton Poth, Aishwarya Kamath, Ivan Vulić, Sebastian Ruder, Kyunghyun Cho, and Iryna Gurevych. 2023. AdapterHub: A Framework for Adapting Transformers. https://adapterhub.ml/. Accessed: 2023-01-22.

[10] Andreas Säuberli, Sarah Ebling, and Martin Volk. 2020. Benchmarking Data-driven Automatic Text Simplification for German. In *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI)*. European Language Resources Association, Marseille, France, 41–48.