Adapting Pre-trained Language Models to Low-Resource Text Simplification: The Path Matters

Cristina Garbacea (garbacea@umich.edu)[1], Qiaozhu Mei[1,2]
[1] EECS Department, [2] School of Information, University of Michigan, Ann Arbor, MI, USA

COLLAs 2022

UNIVERSITY OF MICHIGAN

Background & Motivation

Text simplification aims to reduce the language complexity of highly specialized textual content so that it is accessible for readers who lack adequate literacy skills, such as children, people with low education, people who have reading disorders or dyslexia, and non-native speakers.

Text simplification has considerable potential to improve the fairness and transparency of text information systems, for eg. in healthcare, education, for kids and non-native speakers of a language.

However, text simplification is a low-resource setting. Abundant training examples do not exist or are costly to label, there are only few domains (news, Wikipedia) with parallel-aligned complex-simple examples.

NLP for Low-Resource Settings

NLP solutions for low-resource scenarios:

- **Transfer Learning (Domain Adaptation):**
- source and target domains consist of same feature space (Day, 2017)
- sufficient in-domain data for the target task (Chen et al, 2019)
- Issues: model overfitting, catastrophic forgetting, negative transfer across tasks (Xu et al, 2020), (Thompson et al, 2019)

- **Meta-Learning (Task Adaptation):**

- Promising general learning strategy suitable for few-shot learning and cross-domain generalization (Li et al, 2018), (Wang et al, 2020)
- resource constrained problems where there is a distribution of tasks
- *Metric Learning* (Vinyals et al, 2016), (Snell et al, 2017), *Memory Networks* (Santoro et al, 2016),(Oreshkin et al, 2018), *Gradient based* (Finn et al, 2017), (Zhang et al, 2018)

Research Questions

RQ1: Can we learn to quickly adapt pre-trained language models to new tasks and domains with few training examples in text simplification?

RQ2: Can we combine the advantages of task and domain adaptation?

RQ3: Can consecutive stages of task and domain adaptation improve upon single stage adaptation?

RQ4: When using multiple stages of task and domain adaptation, which is the ideal order in which to perform adaptation?

Proposed Approach

Frame the problem of low-resource text simplification from a **task and domain adaptation perspective**; consider parallel complex-simple examples as samples drawn from a distribution of text generation tasks with **varying constraints on level of text complexity and readability**

Goal: investigate whether transfer learning or meta-learning is a suitable adaptation strategy when there is a distribution of low-resource text simplification tasks/domains

Adapting Pre-trained LMs to Text Simplification

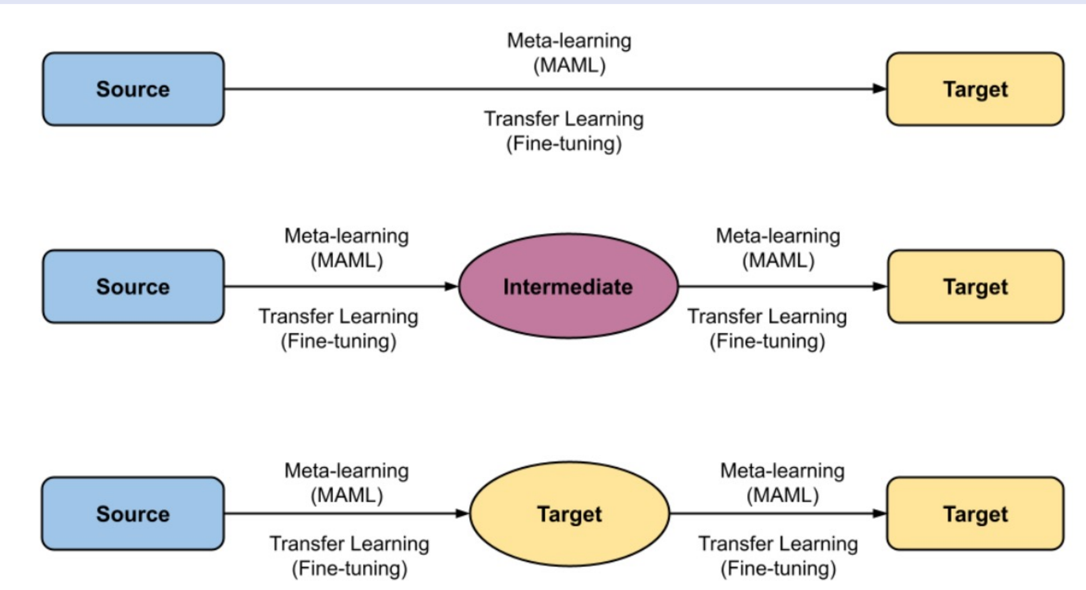

Figure 1: Adapting a pre-trained language model (source) to low-resource text simplification (target).

1. Direct Task Adaptation: determine if it is possible to adapt a pre-trained model (Source) to text simplification tasks (Target) via meta-learning and achieve good performance on unseen tasks (meta-test)

2. Direct Domain Adaptation: establish whether it is possible to adapt a pre-trained model (Source) to text simplification tasks (Target) via fine-tuning and achieve good performance on unseen domains

3. Two-Stage Adaptation via an Intermediate: determine if there is any benefit to combining task and domain adaptation via an intermediate task/domain dataset; find out the order of two-stage adaptation

4. Two-Stage Adaptation via a Pseudostop: if no intermediate dataset is available, determine if Target itself can be used as Intermediate

(Domain Adaptation) Transfer Learning
- Finetune T5 small (60 million parameters)

(Task Adaptation) Gradient-based Meta Learning
- MAML (Finn et al, 2017), Reptile (Nichols et al, 2018)

Experimental Setup

Datasets: different domains, application scenarios of text simplification
- News simplification: ***Newsela*** (Xu et al, 2015)
- Scientific press release: ***Biendata*** (reseach papers -> press release)
- Auxiliary: ***WikiLarge*** (Zhang et al, 2017), ***WikiSmall*** (Zhu et al, 2010)

Baseline (No Adaptation):
Transformer (Vaswani et al, 2017) trained on Newsela and Biendata
Additional Baselines (Pre-trained text simplification models)
ACCESS (Martin et al, 2020), *DMLMTL* (Guo et al, 2018)

Evaluation Metrics: *SARI* (Xu et al, 2016), *BLEU* (Papineni et al, 2002), *FKGL* (Kincaid, 1975), *MoverScore* (Zhao et al, 2019), *MAUVE* (Pillutla et al, 2021), *BARTScore* (Yuan et al, 2021) – Faithfulness, P, R, F1

Research Findings

Baselines:
- ***Transformer performance is suboptimal compared to adaptation*** methods; for Biendata, it oversimplifies scientific content (FKGL < ground-truth FKGL), while when training data is abundant (WikiLarge), neural text simplification performs well without adaptation
- ACCESS and DMLMTL: performance degrades on Biendata
Critical need for task/ domain adaptation!

1. Direct Task Adaptation: MAML T5 > MAML Wiki, Reptile Wiki
Adapting from a powerful pre-trained LM outperforms training a model directly from limited resource

2. Direct Domain Adaptation: T5 > Transformer trained on WikiLarge
Domain adaptation outperforms task adaptation (MAML, Reptile)
Benefit is larger on OOD tasks and domains from Biendata

3. Two-Stage Adaptation via an Intermediate (WikiLarge)
First adapt T5 to WikiLarge, then continue adapting (Newsela/Biendata)
Most promising: ***adapt to new tasks in the first stage*** (MAML, Reptile), then continue ***adapting to new domains in second stage*** (fine-tuning)
Two-stage task and domain adaptation is better than one stage!

4. Two-Stage Adaptation via a Pseudostop (Target)
(T5 → Newsela/Biendata) → Newsela/Biendata
Task + Domain Adaptation remains best adaptation strategy!
Using Target as intermediate dataset overcomes the need for extra data
Higher benefit on OOD tasks and domains from Biendata

Key Takeaways:

Coupling of task and domain adaptation is beneficial!
Adapt to new tasks first, then continue adapting to new domains!
If no intermediate dataset is available, build a Target pseudostop!

This work was in part supported by the National Science Foundation under grant number 1633370.