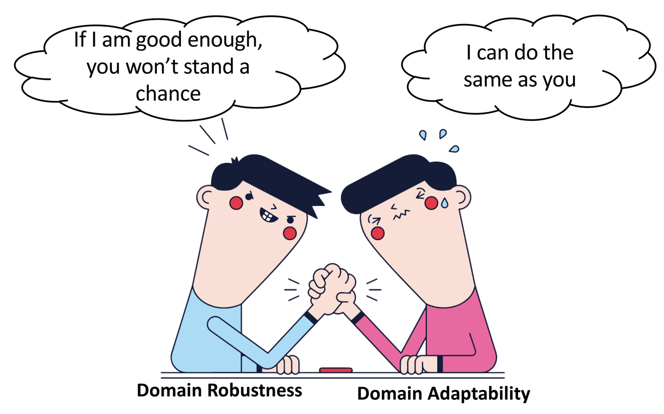
Improving Both Domain Robustness and Domain Adaptability in Machine Translation

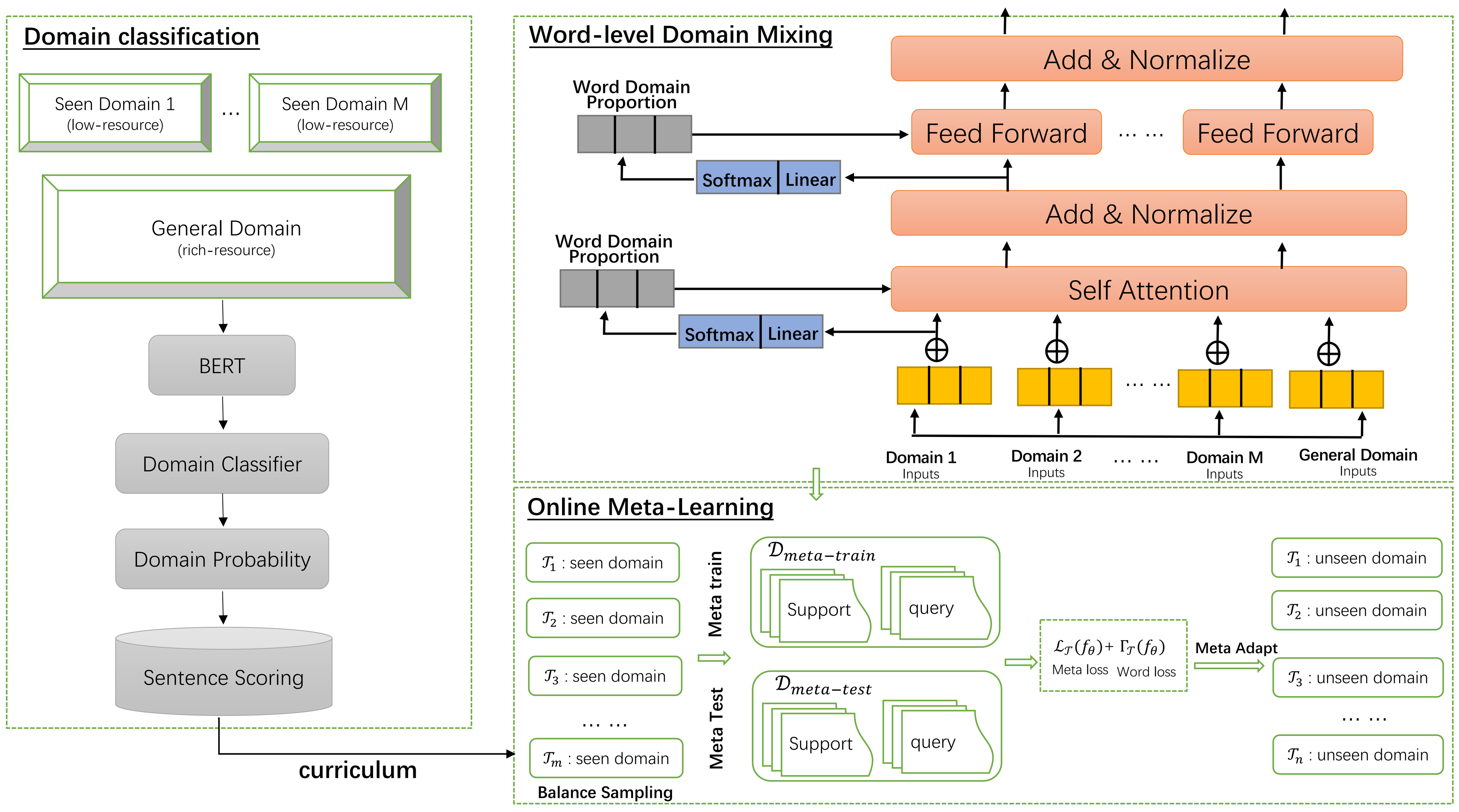
Wen Lai[1], **Jindřich Libovický**[2] and **Alexander Fraser**[1]

[1] Center for Information and Language Processing, LMU Munich, Germany
[2] Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
{lavine, fraser}@cis.lmu.de libovicky@ufal.mff.cuni.cz

Introduction



- We consider two problems of NMT domain adaptation using meta-learning:
  - **Domain Robustness**: we want to reach high quality on both domains seen in the training data and unseen domains.
  - **Domain Adaptability**: making it possible to finetune systems with just hundreds of in-domain parallel sentences.
- We propose a novel approach, **RMLNMT** (**R**obust **M**eta-**L**earning Framework for **N**eural **M**achine **T**ranslation Domain Adaptation), which improves the robustness of existing meta-learning models.



Method

Word-level Domain Mixing

- Motivated by Jiang et al., 2020, we train a word-level domain mixing model, and show that, surprisingly, this improves robustness on unseen domains as well.

Domain Classification

- we train a domain classifier based on BERT (Devlin et al., 2019) to score training sentences; the score measures similarity between out-of-domain and general-domain sentences. This score is used to determine a curriculum to improve the meta-learning process.

Online Meta-Learning

- we improve domain adaptability by integrating the domain-mixing model into a meta-learning framework with the domain classifier using a balanced sampling strategy. The following algorithm shows the complete algorithm.

Algorithm 1 RMLNMT(Robust Meta-Learning NMT Domain Adaptation)
Require: Domain classifier model cls, parameters of pretrained model θ
1: Score the sentence in $\mathcal{D}_{\text{meta-train}}$ using cls
2: **for** N epochs **do**
3: Split corpus into n tasks based on step 1
4: Balance sample through all tasks
5: **for** task $\mathcal{T}_i$, $i = 1 \ldots n$ **do**
6: Evaluate loss $L_{\mathcal{T}}(f_\theta)$
$= \mathcal{L}_{\mathcal{T}_i}(f_\theta) + \Gamma_{\mathcal{T}_i}(f_\theta)$ on support set
7: Update the gradient with parameters
$\theta' = \theta - \alpha \nabla_\theta L_{\mathcal{T}}(f_\theta)$
8: **end for**
9: Update the gradient with parameters
$\theta = \theta - \beta \nabla_\theta L_{\mathcal{T}}(f_{\theta'})$ on query set
10: **end for**
11: **return** RMLNMT model parameter θ

Main Results

Models	Unseen					Seen				
	Covid	Bible	Books	ECB	TED	EMEA	Globalvoices	JRC	KDE	WMT
1 Vanilla	24.34	**12.08**	12.61	29.96	27.89	37.27	24.19	39.84	27.75	27.38
2 Vanilla + tag	24.86	12.04	12.46	30.03	27.93	38.37	24.56	40.75	28.23	27.26
3 Meta-MT w/o FT	23.69	11.07	12.10	29.04	26.86	30.94	23.73	38.82	23.04	26.13
4 Meta-Curriculum (LM) w/o FT	23.70	11.16	12.24	28.22	27.21	33.49	24.27	39.21	27.60	25.83
5 RMLNMT w/o FT	**25.48**	11.48	**13.11**	**31.42**	**28.05**	**47.00**	**26.35**	**51.13**	**32.80**	**28.37**

Table 1: Domain Robustness

Models	Unseen					Seen				
	Covid	Bible	Books	ECB	TED	EMEA	Globalvoices	JRC	KDE	WMT
1 Plain FT	24.81	12.61	12.78	30.48	28.36	37.26	24.26	40.02	27.99	27.31
2 Plain FT + tag	25.31	12.57	12.83	30.57	28.39	39.54	24.91	41.51	29.14	27.58
3 Meta-MT + FT	25.83	14.20	13.39	30.36	28.57	34.69	24.64	39.15	27.47	26.38
4 Meta-Curriculum (LM) + FT	**26.66**	14.37	13.70	30.41	28.97	34.00	24.72	39.61	27.37	26.68
5 RMLNMT + FT	26.53	**15.37**	**13.72**	**31.97**	**29.47**	**47.02**	**26.55**	**51.13**	**32.88**	**28.37**

Table 2: Domain Adaptability

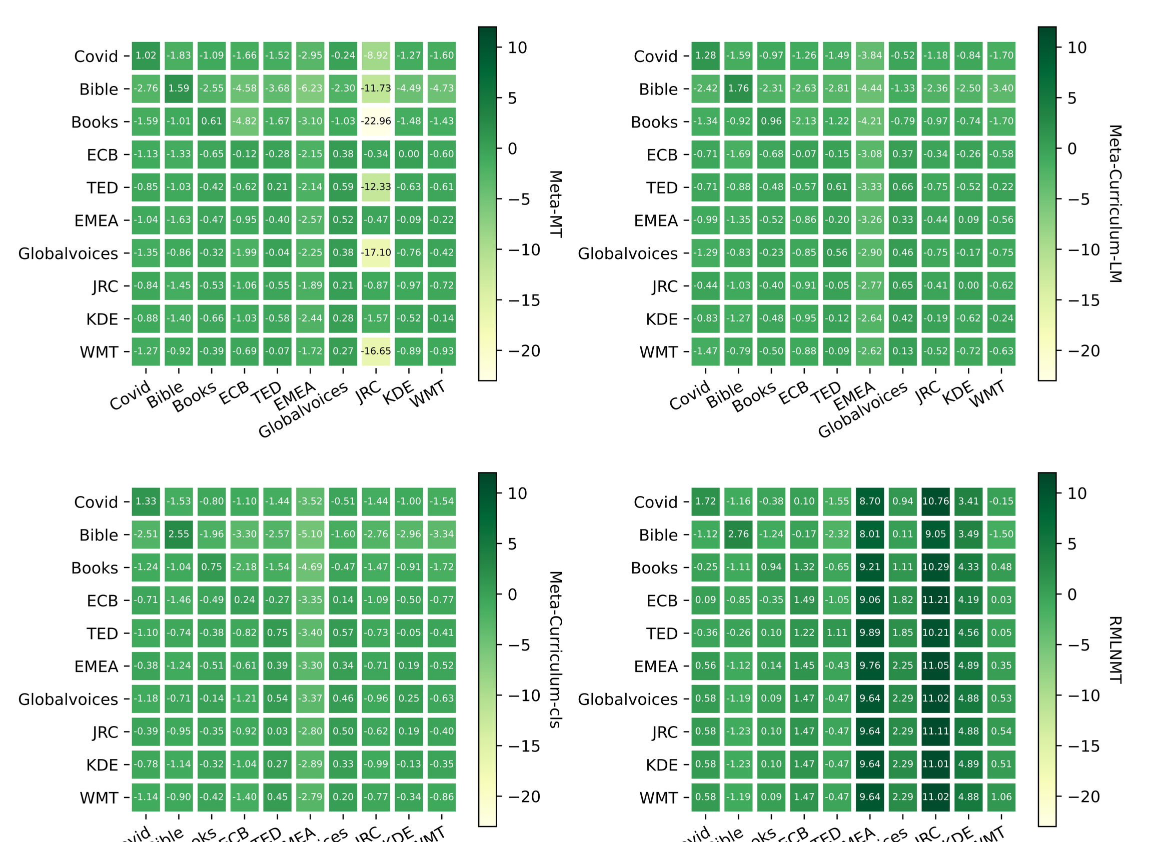


Figure 5: Cross-Domain Robustness

Ablation Study

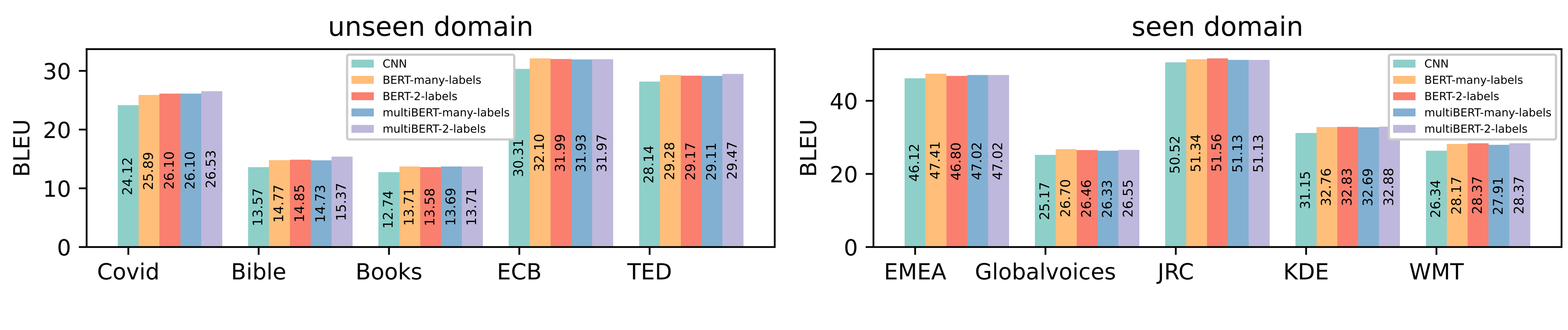


Figure 2: Different Clssifiers

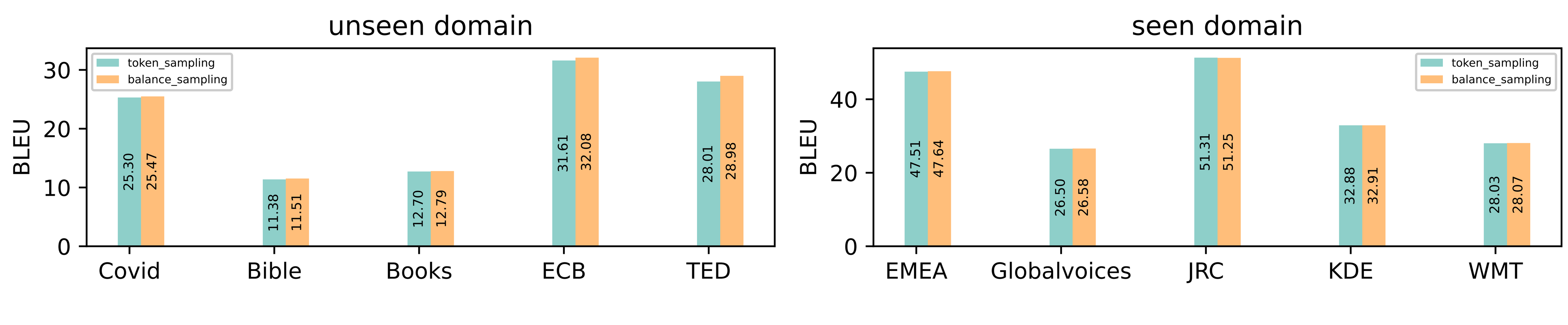


Figure 3: Different Sampling Strategy

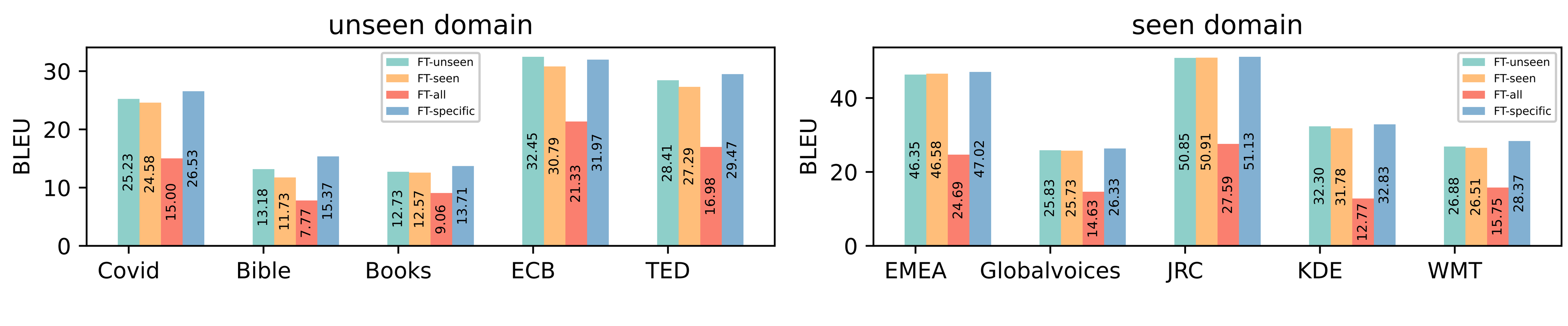


Figure 4: Different Fine-tuning Strategy

Analysis

- **Main Results**
  - **Domain Robustness**. Table 1 shows that RMLNMT got the best domain robustness both in seen and unseen domains.
  - **Domain Adaptability**. From Table 2, we observe that the traditional meta-learning approach shows high adaptability to unseen domains but fails on seen domains due to limited domain robustness. In contrast, RMLNMT shows its domain adaptability both in seen and unseen domains, and maintains the domain robustness simultaneously.
  - **Cross-Domain Robustness**. As shown in Figure 5, Compared with other methods, we observed that RMLNMT shows its robustness on all domains and that the model performance fine-tuned in one specific domain is not sacrificed in other domains.
- **Ablation Study**
  - **Different Classifiers**. In Figure 2, we observed that the performance of RMLNMT is not directly proportional to the accuracy of the classifier.
  - **Different Sampling Strategy**. Figure 3 shows that our methods can result in small improvements in performance.
  - **Different Fine-Tuning Strategy**. We observed on Figure 4 that fine-tuning in one specific domain obtains robust results among all the strategies.

Conclusion

- We presented RMLNMT, a robust meta-learning framework for low-resource NMT domain adaptation reaching both high domain adaptability and domain robustness (both in the seen domains and unseen domains).
- We found that domain robustness dominates the results compared to domain adaptability in meta-learning based approaches.
- The results show that RMLNMT works best in setups that require high robustness in low-resource scenarios.

Acknowledgement

We thank Mengjie Zhao for the helpful comments. This work was supported by funding to Wen Lai's PhD research from LMU-CSC (China Scholarship Council) Scholarship Program. This work has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement #640550). This work was also supported by the DFG (grant FR 2829/4-1). The work at CUNI was supported by the European Commission via its Horizon 2020 research and innovation programme (870930).