Meta-X_{NLG}: A Meta-Learning Approach Based on Language Clustering for Zero-Shot Cross-Lingual Transfer and Generation

Kaushal Kumar Maurya and Maunendra Sankar Desarkar
Indian Institute of Technology Hyderabad, India
cs18resch11003@iith.ac.in & maunendra@cse.iith.ac.in

Introduction

- There are 7000+ spoken languages and 95% of all languages *never* gain attention online today.
- Majority of studies in NLP research are conducted on only English language data [1].
- To democratize the NLP research for the benefit of the large global community, it is *essential* to focus on the non-English languages.
- Recently, cross-lingual transfer learning has emerged as a promising research direction.
- These mechanisms *transfer supervision* from high resource languages like English to low-resource languages or unseen languages like Tamil, Nepali, Swahili, etc.
- Such supervision transfer is *uneven* across languages leading to large performance gaps.
- These performance gaps are observed because models do not account for *cultural* and *linguistic* differences [2].
- This paper is a *step towards* bridging this gap via meta-learning and language clustering.

Intuitive Motivation

- Current active research direction is to learn *shareable structures* across multiple tasks with limited annotated data, for instance, modeling with Meta-Learning algorithms.
- The constraint is: all tasks should share some common structure (or come from a task distribution).
- Different languages in the world follow this constraint as they came into existence with a common goal of communication and share some structure. So, we consider languages as tasks.

Main Objectives

- **Hypothesis:** Meta-learning algorithm trained on typologically diverse languages (as training tasks) can provide *language-agnostic* initialization for the zero-shot cross-lingual generation.
- Towards this- we propose Meta-X_{NLG}, a framework for effective cross-lingual transfer and generation based on *language clustering* and *meta-learning* algorithm.
- To the best of our knowledge -this is the *first* attempt to study meta-learning techniques for cross-lingual natural language generation (X_{NLG}) tasks.
- Particularly, we focus on *zero-shot* X_{NLG} for low-resource languages.
- Among others, we focus on optimization-based meta-learning algorithms, i.e., *Model Agnostic Meta-Learning (MAML)*, due to its recent success.

Methodology

Language Clustering

- Typologically learned (with linguistic features from WALS and URIEL databases) and task-learned (e.g. language tag from MT) language representations are combined using singular vector canonical correlation (SVCC) to obtain *multi-view* language representation [3].
- We use this multi-view language representation to cluster the languages and find representative (i.e., *centroid*) languages.
- Formally, given a cluster $C = \{L_1, L_2, \ldots L_t\}$, where each L_i is multi-view representation of i^{th} language and distance function (e.g. cosine distance; d), the centroid language $L^* \in C$ is defined as:

$$L^* = \arg\min_{L_i \in C} \sum_{L_j \in C} d(L_j, L_i) \quad (1)$$

- Centroid languages are considered as Meta-train languages and Non-centroid languages are considered as Target (Meta-test) languages.

Model Training

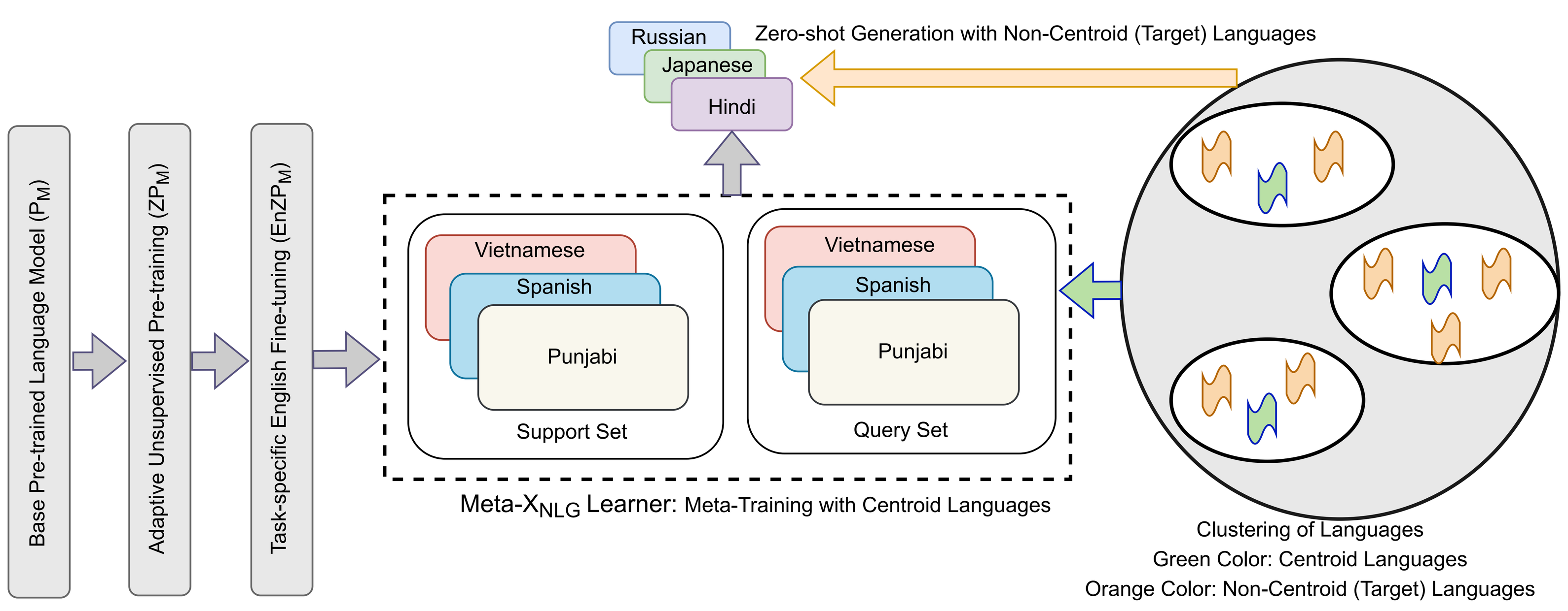

Figure 1: Overview of Meta-X_{NLG} Framework

.

Step-1: Selection of base Pre-trained model (e.g. mT5)
Step-2: *Adaptive* Unsupervised Pre-training to avoid Accidental Translation (AT) problem
Step-3: Fine-tuning on task-specific High-resource language i.e., English
Step-4: *Meta-Training* with Low-resource centroid languages
Step-5: *Meta-adaptation* with Low-resource non-centroid languages in zero-shot setting

Avoiding Accidental Translation

In the zero-shot setting, models suffer in well-formed generation for unseen low-resource languages. The problem is known as Accidental Translation problem. We use following steps to mitigate this:
1. Adding Language Tag: $< fxx >< 2xx >$
2. Adaptive unsupervised pre-training
3. *Freezing model Components:* Freeze token embedding and all decoder layers parameters

Experimental Setup and Results

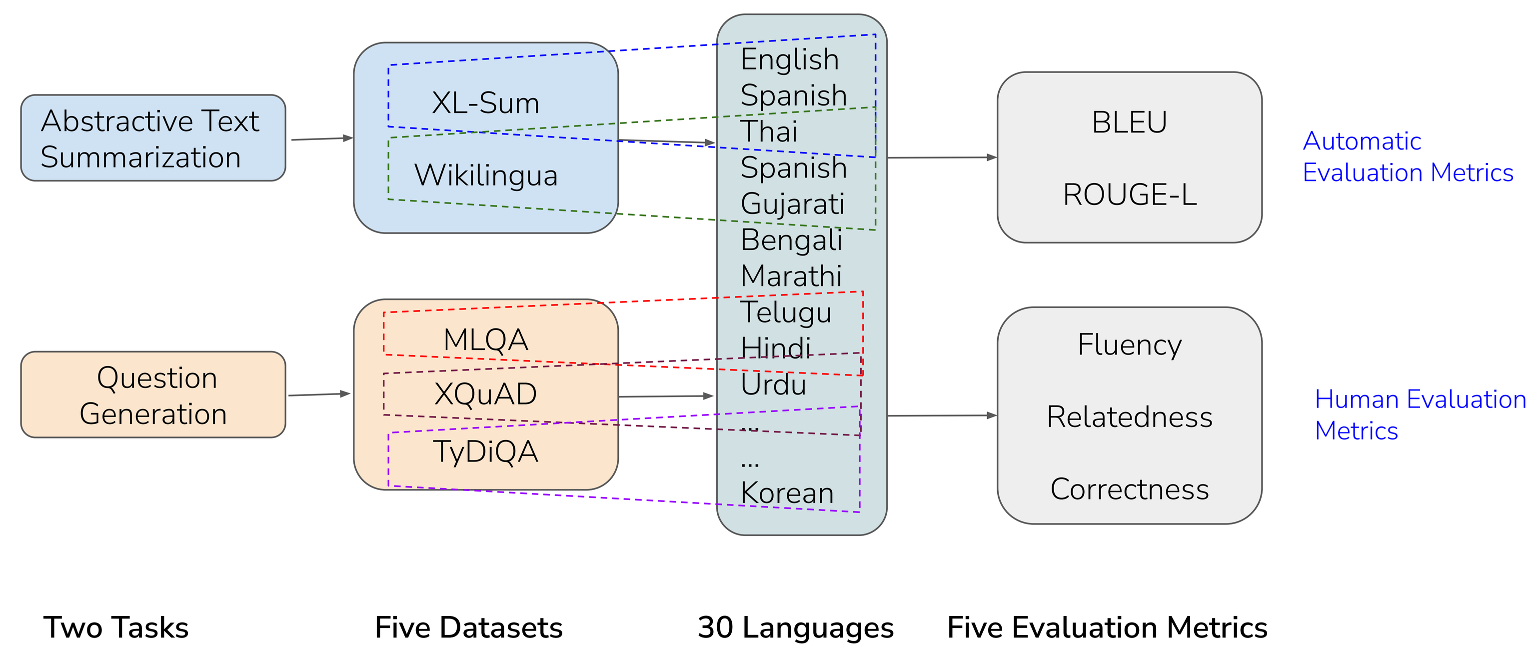

Figure 2: Experimental Setup

Table-1 shows the human evaluation results. The results for automated evaluation are not included here due to space limitations. Those are included and analyzed in detail in the main paper.

Model	Task/Data/Lang	Flu	Rel	Corr	Task/Data/Lang	Flu	Rel	Corr
Annotator set-1								
EnZmT5		4.06	3.58	2.84		4.28	3.94	3.70
FTZmT5	ATS/XL-Sum/bn	2.82	3.18	2.08	ATS/XL-Sum/te	3.46	3.46	3.22
Meta-X_{NLG}		**4.12**	**4.34**	**3.44**		**4.50**	**4.22**	**4.04**
Annotator set-2								
EnZmT5		3.70	3.23	3.26		3.56	3.50	3.20
FTZmT5	ATS/XL-Sum/bn	2.62	2.48	2.16	ATS/XL-Sum/te	3.02	2.84	2.60
Meta-X_{NLG}		**3.97**	**3.48**	**3.28**		**4.18**	**4.10**	**3.88**
Annotator set-1								
EnZmT5		4.00	3.72	3.68		4.12	4.24	2.54
FTZmT5	ATS/Wiki/hi	4.07	3.39	3.83	QG/XQuAD/hi	4.22	4.02	2.56
Meta-X_{NLG}		**4.09**	**3.80**	**3.97**		**4.42**	**4.34**	**2.86**
Annotator set-2								
EnZmT5		4.38	4.22	4.00		3.28	3.63	2.82
FTZmT5	ATS/Wiki/hi	4.57	**4.44**	4.08	QG/XQuAD/hi	3.24	3.34	2.89
Meta-X_{NLG}		**4.66**	**4.44**	**4.16**		**3.59**	**3.67**	**3.24**
Annotator set-1								
EnZmT5		3.48	3.70	3.46		4.25	4.06	3.10
FTZmT5	QG/MLQA/hi	3.44	3.42	3.18	QG/TyDiQA/ta	3.25	3.01	2.07
Meta-X_{NLG}		**3.70**	**3.74**	**3.56**		**4.74**	**4.20**	**3.39**
Annotator set-2								
EnZmT5		**3.30**	3.28	2.40		3.00	4.08	2.82
FTZmT5	QG/MLQA/hi	3.10	3.44	2.84	QG/TyDiQA/ta	2.55	3.045	1.83
Meta-X_{NLG}		3.24	**3.70**	**2.88**		**4.04**	**4.46**	**3.20**

Table 1: Human Evaluation results for four languages (**hi:** Hindi, **te:** Telugu, **ta:** Tamil and **bn:** Bengali), two annotator sets, two tasks (**ATS** and **QG**) and all five datasets. **Flu:** Fluency, **Rel:** Relatedness and **Corr:** Correctness metrics. Results are shown for two annotation sets which ensure biased free evaluation. Reported scores are average of all the annotators in a annotator set. All metrics are evaluated at scale of 1-5 where 1 is low and 5 is high.

Conclusions

- We propose a novel Meta-X_{NLG} framework based on meta-learning and language clustering for effective cross-lingual transfer and generation.
- This is the first study that uses meta-learning for zero-shot cross-lingual transfer and generation.
- The evaluations are done on two challenging tasks (ATS and QG), five publicly available datasets and 30 languages and consistent improvements are observed.
- In the future, we will extend this study to more cross-lingual tasks and languages.

References

[1] Emily M Bender. The# benderrule: On naming the languages we study and why it matters. *The Gradient*, 14, 2019.

[2] Damián Blasi, Antonios Anastasopoulos, and Graham Neubig. Systematic inequalities in language technology performance across the world's languages. *arXiv preprint arXiv:2110.06733*, 2021.

[3] Arturo Oncevay, Barry Haddow, and Alexandra Birch. Bridging linguistic typology and multilingual machine translation with multi-view language representations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2391–2406, Online, November 2020. Association for Computational Linguistics.

Acknowledgements

We thank all the human evaluators for evaluation, anonymous reviewers for constructive feedback. We also thank Microsoft and ACM-India/IARCS for the travel grant to attend the conference.