NATURAL LANGUAGE
& INFORMATION
PROCESSING LAB

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

ACL 2022
22ND – 27TH MAY | 60TH MEETING | DUBLIN

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

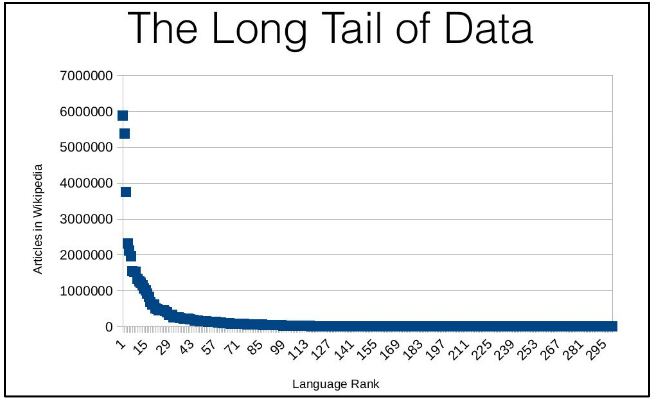
Kaushal Kumar Maurya and Maunendra Sankar Desarkar,
Department of Computer Science and Engineering
Indian Institute of Technology Hyderabad, India

ACL 2022(Findings)

Outline

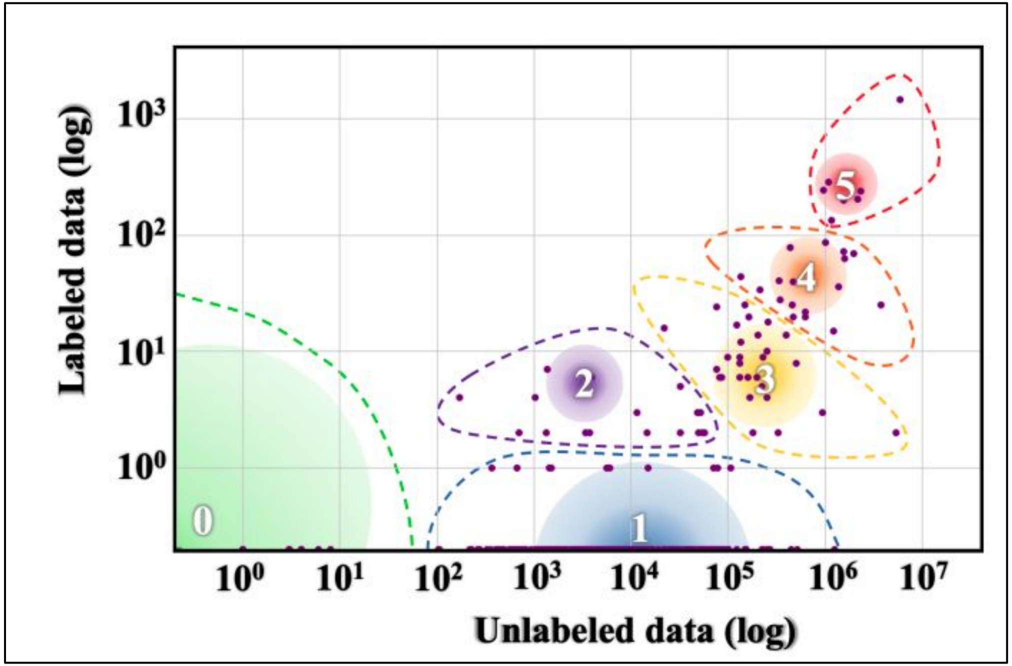
- ❑ Introduction
- ❑ Background: MAML
- ❑ Proposed Approach
- ❑ Experiment Setup
- ❑ Results and Analysis
- ❑ Conclusion
- ❑ Other Project: MT for ELRLs

Introduction: Need of Multilingual Models


The Long Tail of Data

Most Spoken Languages of the World

1. English (1.132 B)
2. 中文(普通话) (1.116 B)
3. हिन्दी (615.4 M)
4. Español (534.4 M)
5. Français (279.8 M)
6. اَلْعَرَبِيَّةُ (273.9 M)
7. বাংলা (265.0 M)
8. Росси́я (258.2 M)
9. Português (234.1 M)
10. Bahasa Indonesia (279.8 M)

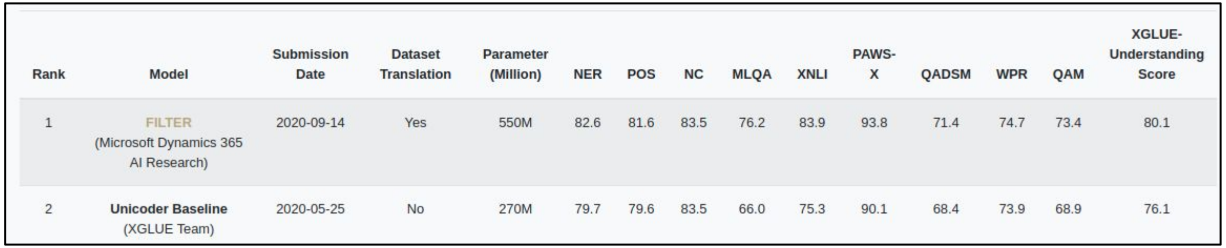


Source: http://langscape.umd.edu/map.php

- 7000+ spoken languages (https://www.ethnologue.com/)
- 95% of all languages in use today will never gain attraction online (Andras Kornai)

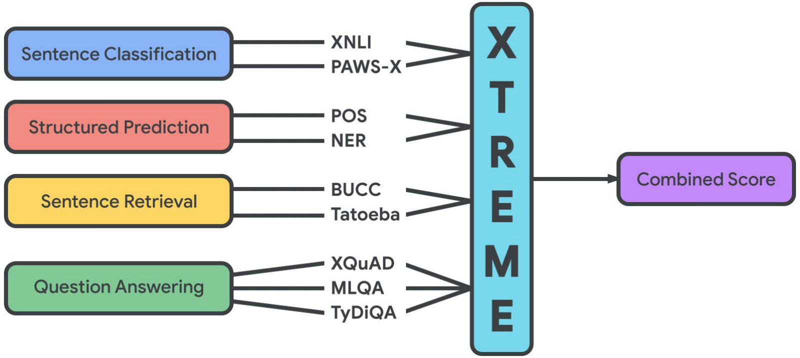
Reference: Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 6282–6293, Online. Association for Computational Linguistics.

Introduction: Efforts in Cross-lingual Research



Rank	Model	Submission Date	Dataset Translation	Parameter (Million)	NER	POS	NC	MLQA	XNLI	PAWS-X	QADSM	WPR	QAM	XGLUE-Understanding Score
1	FILTER (Microsoft Dynamics 365 AI Research)	2020-09-14	Yes	550M	82.6	81.6	83.5	76.2	83.9	93.8	71.4	74.7	73.4	80.1
2	**Unicoder Baseline** (XGLUE Team)	2020-05-25	No	270M	79.7	79.6	83.5	66.0	75.3	90.1	68.4	73.9	68.9	76.1

XGLUE-Benchmark [1]



XTREME-Benchmark [2]

1. Supervision Transfer from high resource languages (mainly English) to low resource languages
2. Modeling is mostly based on limited supervision setting (zero-shot and few-shot setting)
3. Covers large set of languages and multiple tasks
4. New modeling techniques are developing every day

Reference: [1] Liang, Yaobo, et al. "Xglue: A new benchmark dataset for cross-lingual pre-training, understanding and generation." *arXiv preprint arXiv:2004.01401* (2020).
[2] Hu, Junjie, et al. "Xtreme: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation." *International Conference on Machine Learning*. PMLR, 2020.

Introduction: Limitations and Hope

Limitations:

- ➢ Supervision transfer is uneven across languages
  - ○ Leads to large performance gaps.
- ➢ Cultural and linguistic differences are not considered in the modeling [3,4]

Hope:

- ➢ One active research direction is to learn shareable structures across multiple tasks with limited annotated data (Modeling with Meta-Learning algorithms)
  - ○ Constraint: all tasks should share some common structure (or come from a task distribution)
- ➢ The world's different languages follow this constraint
  - ○ Came into existence with a common goal of communication, and share some structure.
- ➢ We consider all the languages as tasks

Reference: [3] Guokun Lai, Barlas Oguz, Yiming Yang, and Veselin Stoyanov. 2019. Bridging the domain gap in cross-lingual document classification. CoRR, abs/1909.07009.
[4] Damián Blasi, Antonios Anastasopoulos, and Graham Neubig. 2021. Systematic inequalities in language technology performance across the world's languages. arXiv preprint arXiv:2110.06733.

Introduction

Hypothesis: Meta-learning algorithm trained on *typologically diverse* languages (as training task) provide language-agnostic initialization for the zero-shot cross-lingual generation.

- ➢ We propose Meta-X_{NLG}, a framework for effective cross-lingual transfer and generation based on Model-Agnostic Meta-Learning (MAML) algorithm.
- ➢ This is first attempt to study meta-learning techniques for cross-lingual natural language generation (X_{NLG}).
- ➢ Particularly, we focus on zero-shot X_{NLG} for low-resource languages.

Outline

- ❏ Introduction
- ❏ Background: MAML
- ❏ Proposed Approach
- ❏ Experiment Setup
- ❏ Results and Analysis
- ❏ Conclusion
- ❏ Other Project: MT for ELRLs

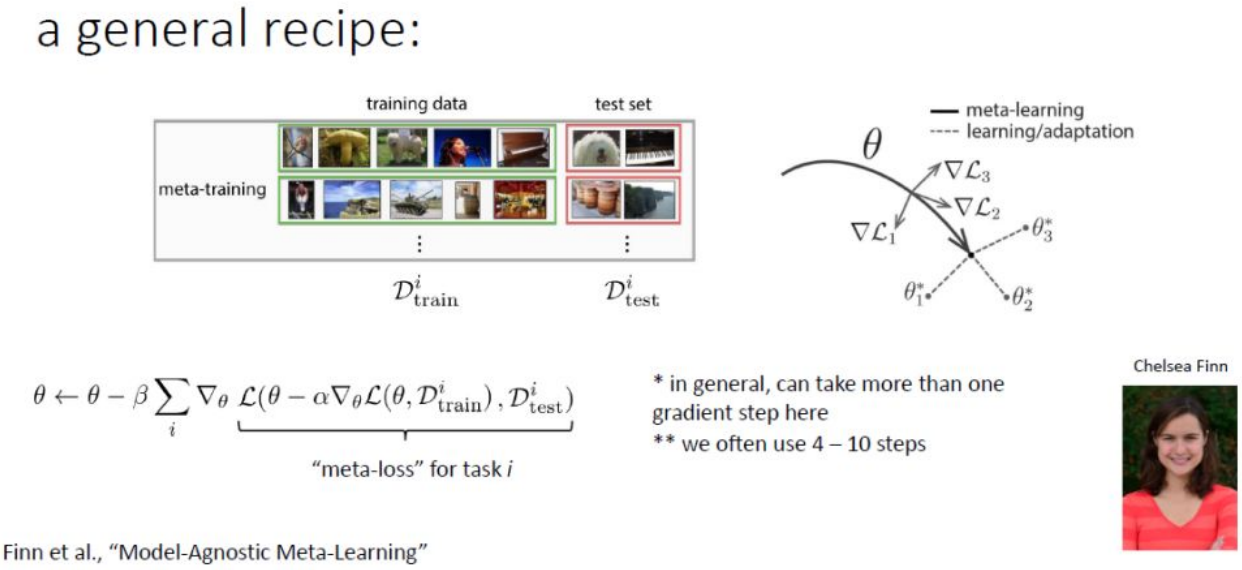
Model-Agnostic Meta-Learning (MAML)

- Meta-learning or learning to learn[1]
  - A learning paradigm
  - Model is trained on diverse tasks (and learns structures)
  - Quickly adapts to new tasks given a handful of examples.

- Among others, we focus on optimization-based algorithms, i.e., Model Agnostic Meta-Learning (MAML) due to its recent success.

Reference: [1] Yoshua Bengio, Samy Bengio, and Jocelyn Cloutier. 1990. Learning a synaptic learning rule. Citeseer.

Model-Agnostic Meta-Learning (MAML)



a general recipe:

$$\theta \leftarrow \theta - \beta \sum_i \nabla_\theta \underbrace{\mathcal{L}(\theta - \alpha \nabla_\theta \mathcal{L}(\theta, \mathcal{D}^i_{\text{train}}), \mathcal{D}^i_{\text{test}})}_{\text{“meta-loss” for task } i}$$

* in general, can take more than one gradient step here

** we often use 4 – 10 steps

Chelsea Finn

Finn et al., “Model-Agnostic Meta-Learning”

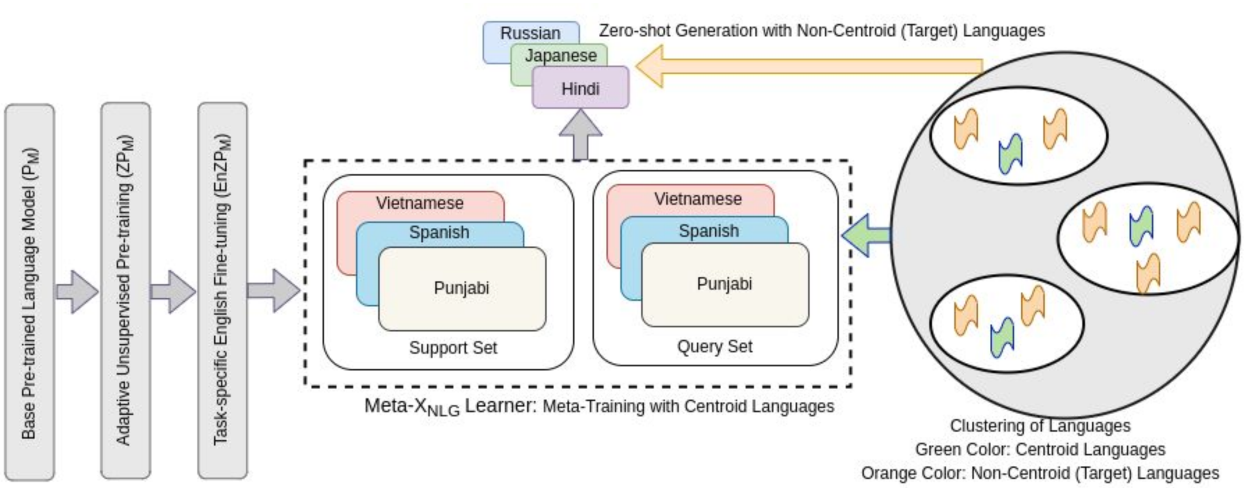
Slide Credit: Sergey Levine

Reference: [5] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In International Conference on Machine Learning, pages 1126–1135. PMLR.

Outline

- ❑ Introduction
- ❑ Background: MAML
- ❑ Proposed Approach
- ❑ Experiment Setup
- ❑ Results and Analysis
- ❑ Conclusion
- ❑ Other Project: MT for ELRLs

Proposed Approach: Summary



1. Cluster the languages and find the centroid languages
2. Train MAML with centariod languages
3. Evaluate the Trained MAML model with non-centroid languages in zero-shot setting while overcoming accidental translation problem

Proposed Approach: Intuition

- **Generalization Goals:** Intra-cluster and Intra-cluster
    - **Intra-cluster:** Training with one central language (centroid) per cluster
    - **Inter-cluster:** Training with multiple centroid languages
- **Number of clusters (meta-training language) Vs Generalization**
    - Single cluster => over-generalization [fails to learn different typological structures]
    - Too many clusters => Many centroid languages => Many typological structures => learning distracted
    - Empirical evidence => Three clusters provide best Generalization

Proposed Approach: Language Clustering

Representation of Languages:

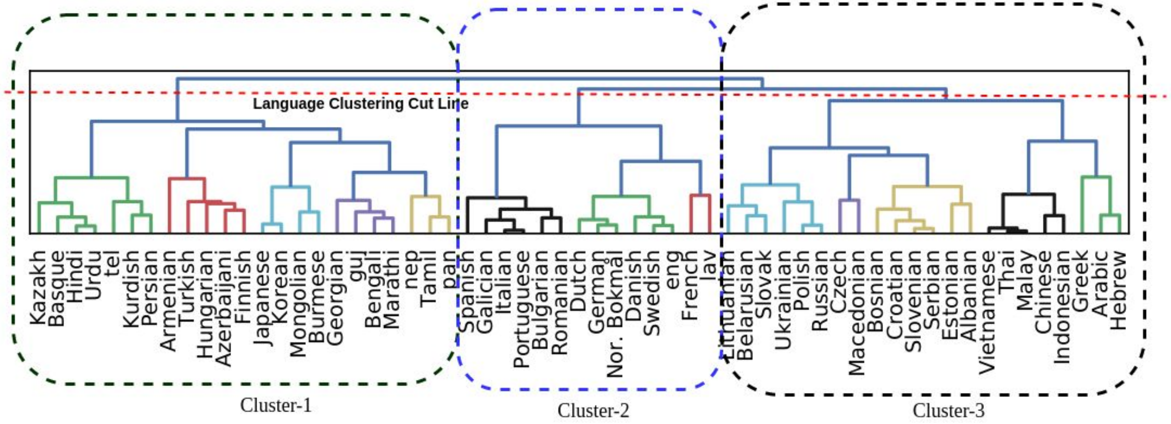
1. **Typologically Learned:** Use typological information from linguistic knowledge-base like WALS [6]
2. **Task-learned:** Extract learned language tag representations from tasks like machine translation [7]
3. **Multi-View:** Fuse typologically learned and task-learned language representations using singular vector canonical correlation (SVCC) as proposed by Oncevay et al. (2020) [8]

Reference: [6] Matthew S Dryer and Martin Haspelmath. 2013. The world atlas of language structures online

[7] Chaitanya Malaviya, Graham Neubig, and Patrick Littell. 2017. Learning language representations for typology prediction. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 2529–2535, Copenhagen, Denmark. Association for Computational Linguistics

[8] Arturo Oncevay, Barry Haddow, and Alexandra Birch. 2020. 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. Association for Computational Linguistics.

Proposed Approach: Language Clustering



Language

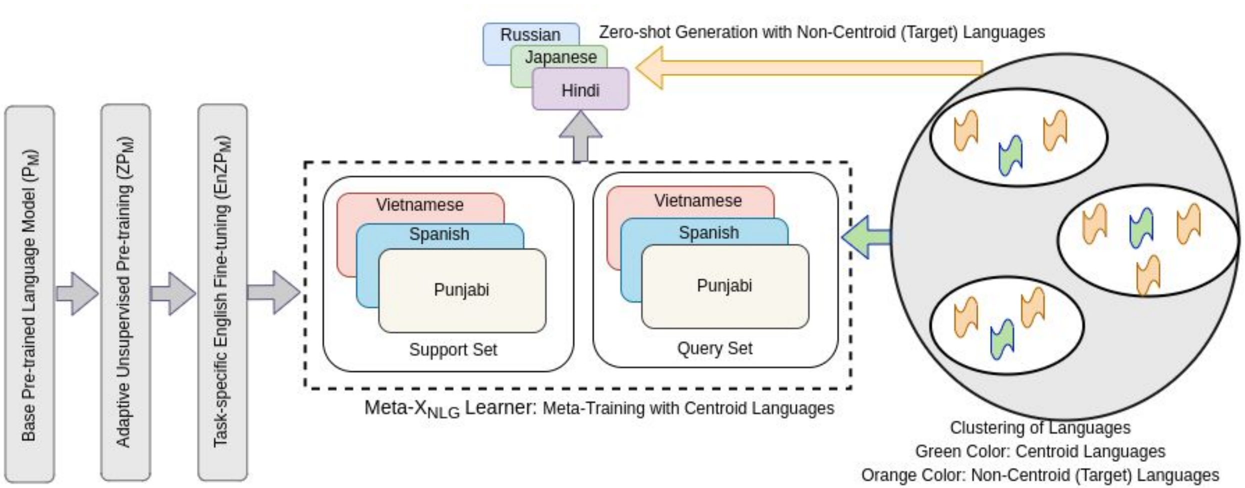
Clustering

- Given cluster $C = \{L_1, L_2, \ldots L_t\}$ and d is cosine distance then the centroid language $L^* \in C$ is defined as:

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

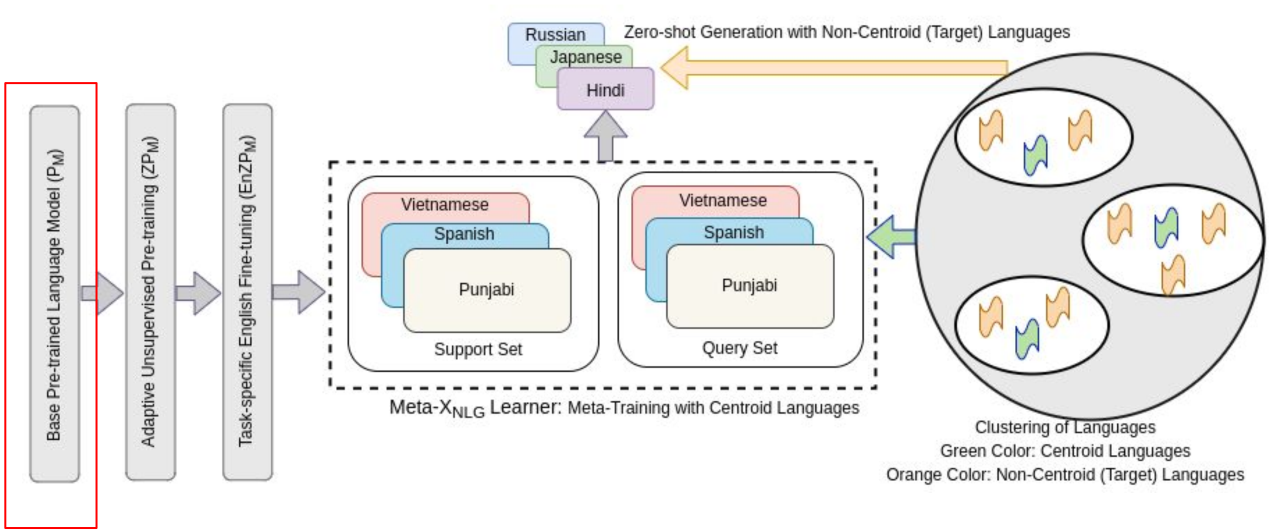
Cluster-1(14)	**Cluster-2(8)**	**Cluster-3(8)**
hi,ur,te,tr,ja,fi,ko,gu,bn,mr,np,ta,pa,sw	es,it,pt,ro,nl,de,en,fr	ru,cs,vi,th,zh,id,el,ar

Proposed Approach: Model Training



1. [Step-01] Selection of Base Pre-trained Model
2. [Step-02] Adaptive Unsupervised Pre-training
3. [Step-03] Fine-tuning on High Resource Language
4. [Step-04] Meta-Training with Low-resource Centroid Languages
5. [Step-05] Meta-adaptation for Zero-shot Evaluation

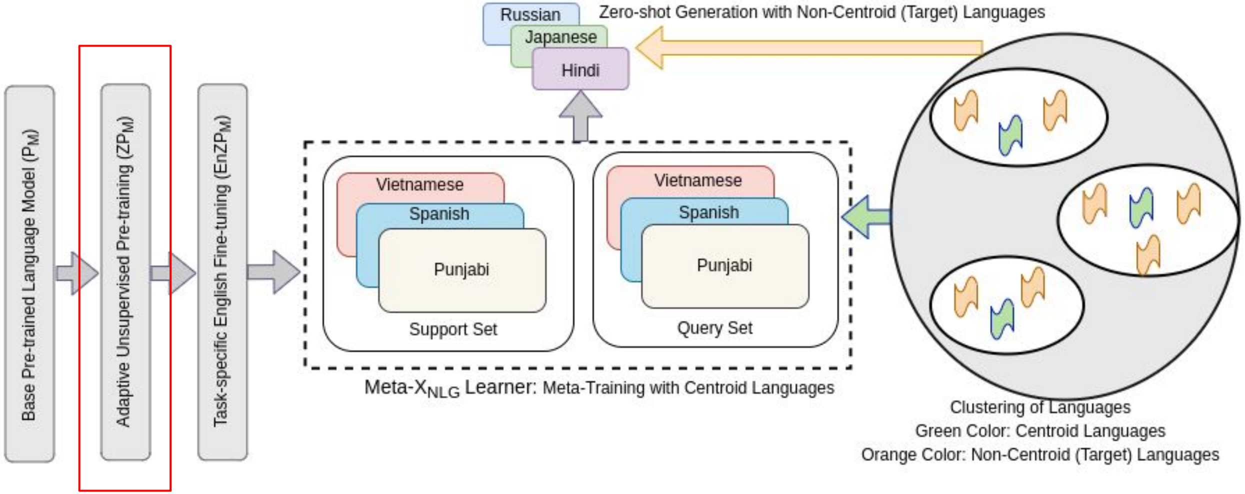
Proposed Approach: Model Training



[Step-01] Selection of Base Pre-trained Model:

1. Model-agnostic
2. Use sequence-to-sequence multilingual pre-trained language
3. We use mT5

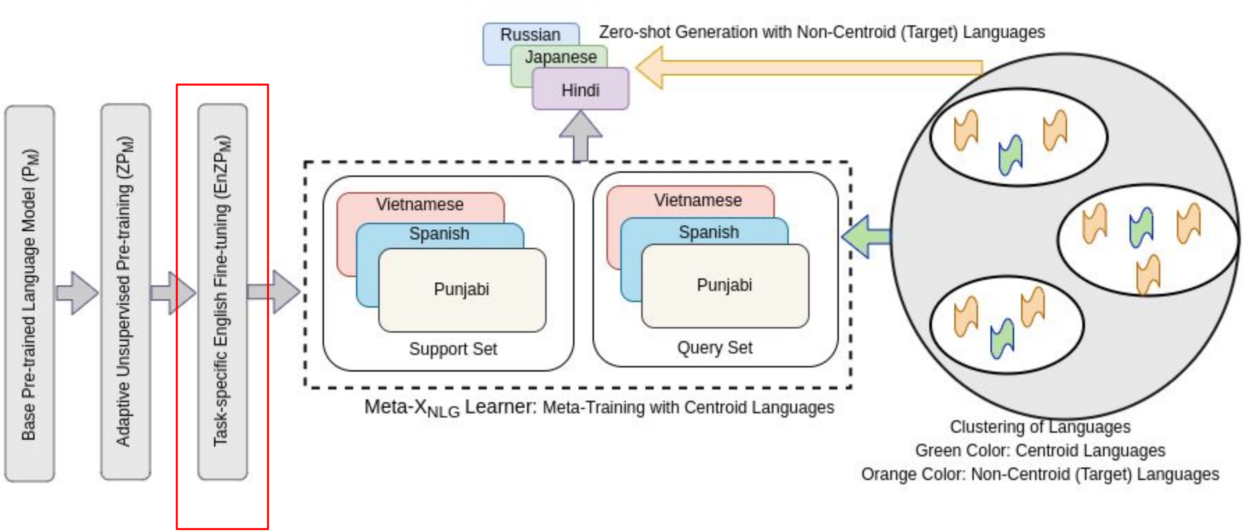
Proposed Approach: Model Training



[Step-02] Adaptive Unsupervised Pre-training:

1. An adaptive pre-training step on top of mT5
2. Use mT5 denoising objective
3. Mitigate Accidental Translation problem

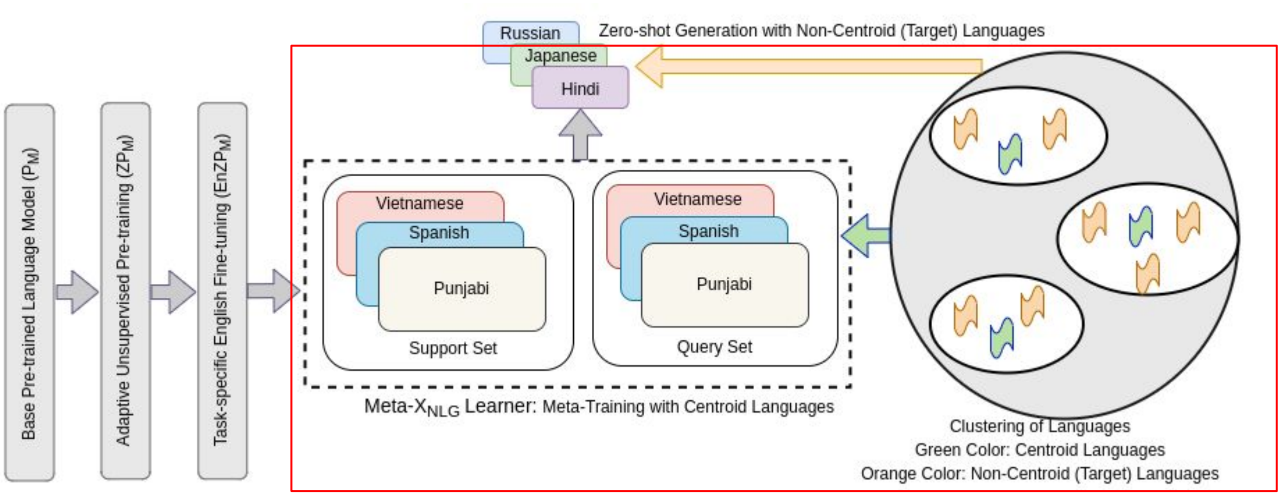
Proposed Approach: Model Training



[Step-03] Fine-tuning on High Resource Language:

1. Task specific training with English dataset
2. The supervision will transfer from English

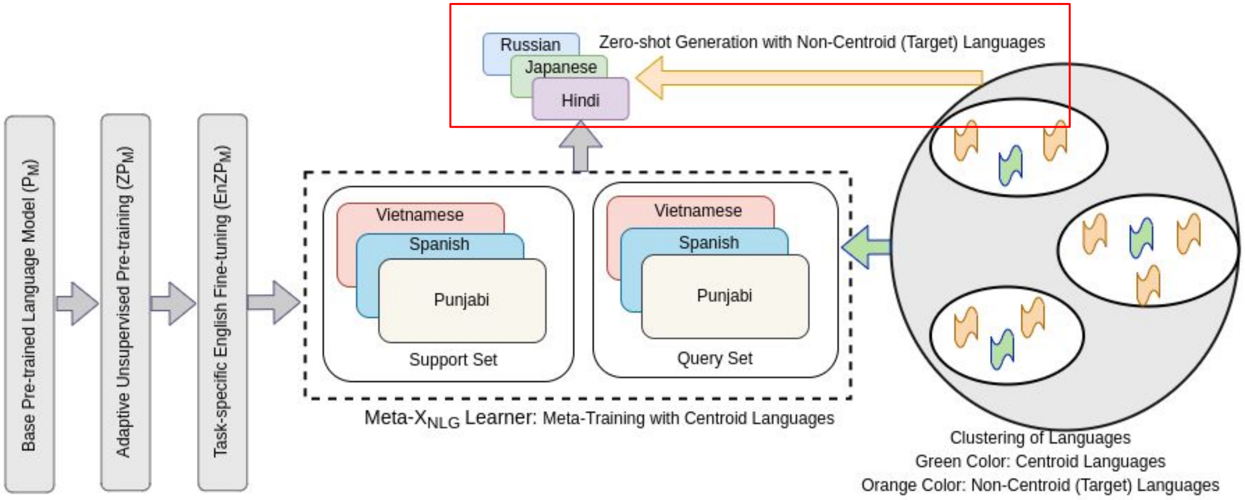
Proposed Approach: Model Training



[Step-04] Meta-Training with Low-resource Centroid Languages:

1. Train MAML with centroid languages
2. **Obtain meta-learned checkpoints**

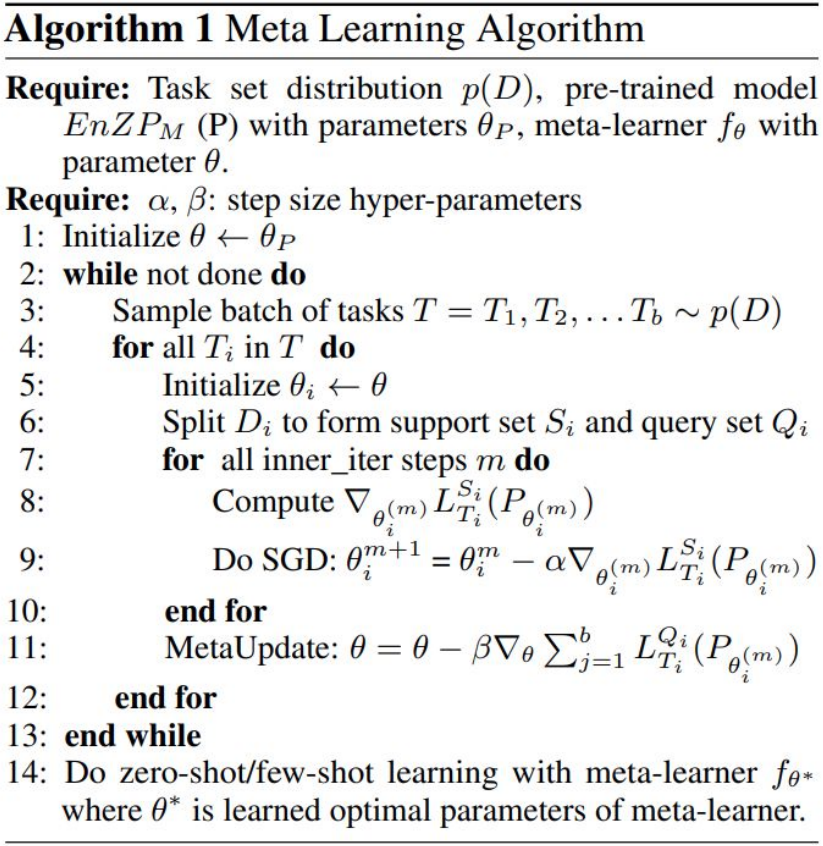
Proposed Approach: Model Training



[Step-05] Meta-adaptation for Zero-shot Evaluation:

1. Evaluate the meta-learned model with non-centroid languages
2. Evaluation is done in Zero-shot setting

Proposed Approach: Algorithm



Algorithm 1 Meta Learning Algorithm

```
Require: Task set distribution $p(D)$, pre-trained model $EnZP_M$ (P) with parameters $\theta_P$, meta-learner $f_\theta$ with parameter $\theta$.
Require: $\alpha$, $\beta$: step size hyper-parameters
Initialize $\theta \leftarrow \theta_P$
while not done do
    Sample batch of tasks $T = T_1, T_2, \dots T_b \sim p(D)$
    for all $T_i$ in $T$ do
        Initialize $\theta_i \leftarrow \theta$
        Split $D_i$ to form support set $S_i$ and query set $Q_i$
        for all inner_iter steps $m$ do
            Compute $\nabla_{\theta_i^{(m)}} L_{T_i}^{S_i}(P_{\theta_i^{(m)}})$
            Do SGD: $\theta_i^{m+1} = \theta_i^m - \alpha \nabla_{\theta_i^{(m)}} L_{T_i}^{S_i}(P_{\theta_i^{(m)}})$
        end for
        MetaUpdate: $\theta = \theta - \beta \nabla_\theta \sum_{j=1}^{b} L_{T_i}^{Q_i}(P_{\theta_i^{(m)}})$
    end for
end while
Do zero-shot/few-shot learning with meta-learner $f_{\theta^*}$ where $\theta^*$ is learned optimal parameters of meta-learner.
```

Avoiding Accidental Translations

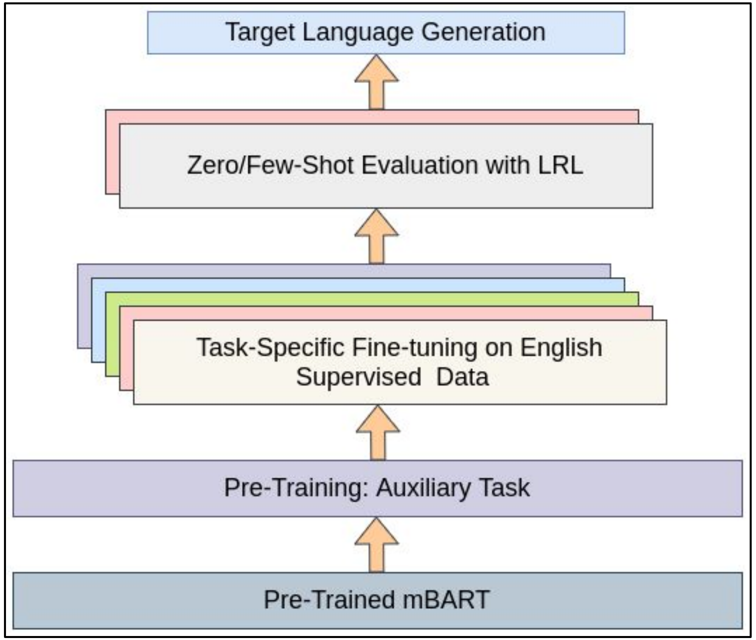
- In zero-shot setting, model suffers from ill-formed generation for unseen low-resource and problem is known as Accidental Translation (AT)
- Where model generates whole/part of the output in the fine-tuning language [9]
- Aligned with catastrophic forgetting problem

Reference: [9] Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 483–498, Online. Association for Computational Linguistics.

ZmBART Model: ACL 2021

ZmBART: An Unsupervised Cross-lingual Transfer Framework for Language Generation

- ➢ Proposed unsupervised *framework (ZmBART)* to mitigate accidental translation/ Catastrophic Forgetting problem
- ➢ Enable zero-shot and few-shot target language generation
- ➢ Leverage mBART pre-trained Model.
- ➢ Evaluated with 4 task and 3 languages

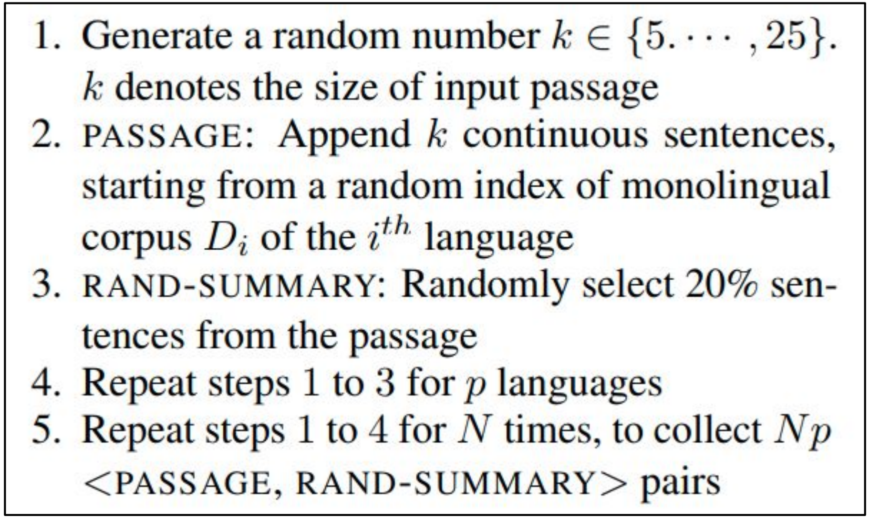


Reference: [10] Kaushal Kumar Maurya, Maunendra Sankar Desarkar, Yoshinobu Kano, and Kumari Deepshikha. 2021. Zmbart: An unsupervised cross-lingual transfer framework for language generation. In Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021, volume ACL/IJCNLP 2021 of Findings of ACL, pages 2804–2818. Association for Computational Linguistics.

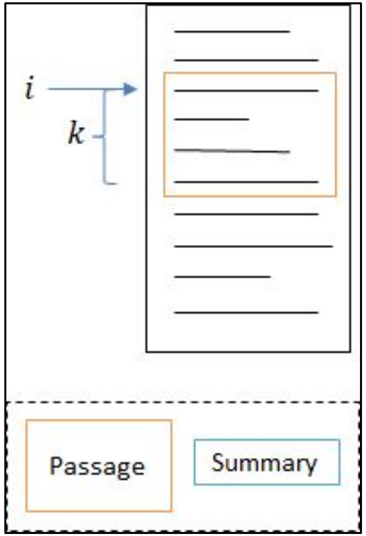
Auxiliary Task: Adaptive Pre-training

“Given an input passage, generate few random sentences (called rand-summary) from the passage”

Data preparation steps for the auxiliary task are given below:



1. Generate a random number $k \in \{5. \cdots, 25\}$. k denotes the size of input passage
2. PASSAGE: Append k continuous sentences, starting from a random index of monolingual corpus D_i of the i^{th} language
3. RAND-SUMMARY: Randomly select 20% sentences from the passage
4. Repeat steps 1 to 3 for p languages
5. Repeat steps 1 to 4 for N times, to collect Np <PASSAGE, RAND-SUMMARY> pairs



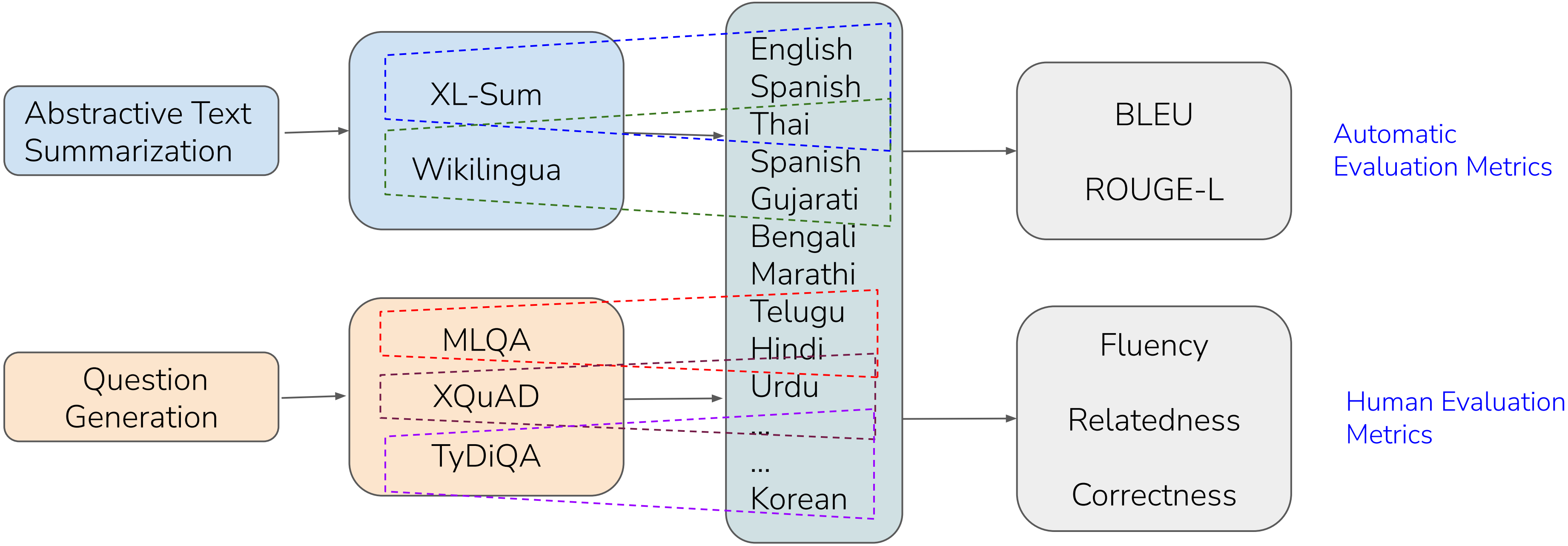
What Worked?

1. Added Language Tag: <fxx><2xx>
2. Adaptive Unsupervised Pre-training
3. Freezing model Components: Freeze token embedding and all the parameters of decoder layers

Outline

- ❑ Introduction
- ❑ Background: MAML
- ❑ Proposed Approach
- ❑ Experimental Setup
- ❑ Results and Analysis
- ❑ Conclusion
- ❑ Other Project: MT for ELRLs

Experimental Setup



Abstractive Text Summarization

XL-Sum

Wikilingua

Question Generation

MLQA

XQuAD

TyDiQA

English
Spanish
Thai
Spanish
Gujarati
Bengali
Marathi
Telugu
Hindi
Urdu
...
...
Korean

BLEU

ROUGE-L

Automatic Evaluation Metrics

Fluency

Relatedness

Correctness

Human Evaluation Metrics

Two Tasks

Five Datasets

30 Languages

Five Evaluation Metrics

Baselines

- **EnZmT5:** Inspired from [10], Direct fine-tuning + adaptive step
- **FTZmT5:** Inspired from [11], fine-tune EnZmT5 with centroid languages.

Reference: [10] Kaushal Kumar Maurya, Maunendra Sankar Desarkar, Yoshinobu Kano, and Kumari Deepshikha. 2021. Zmbart: An unsupervised cross-lingual transfer framework for language generation. In Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021, volume ACL/IJCNLP 2021 of Findings of ACL, pages 2804–2818. Association for Computational Linguistics

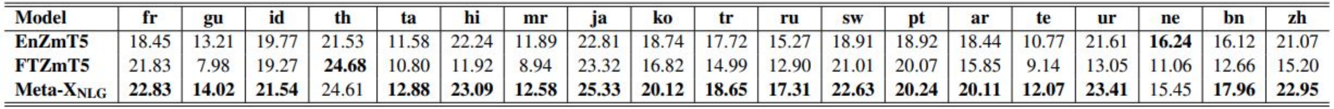
[11] Mike Lewis, Marjan Ghazvininejad, Gargi Ghosh, Armen Aghajanyan, Sida Wang, and Luke Zettlemoyer. 2020a. Pre-training via paraphrasing.

Outline

- ❏ Introduction
- ❏ Background: MAML
- ❏ Proposed Approach
- ❏ Experiment Setup
- ❏ Results Analysis
- ❏ Conclusion
- ❏ Other Project: MT for ELRLs

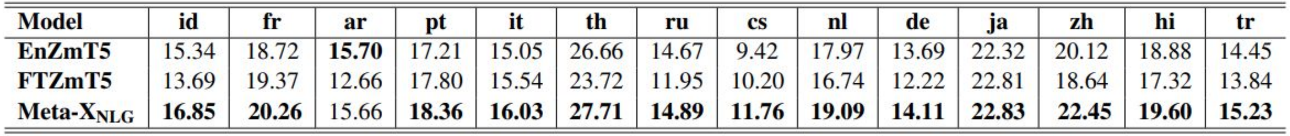
Results: Automatic Evaluation Scores

XL-SUM



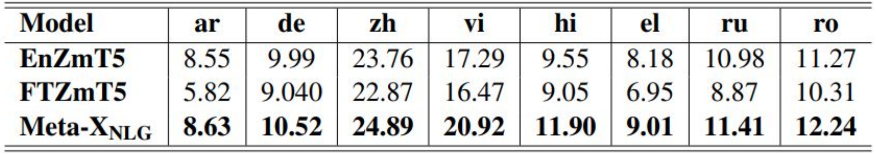
Model	fr	gu	id	th	ta	hi	mr	ja	ko	tr	ru	sw	pt	ar	te	ur	ne	bn	zh
EnZmT5	18.45	13.21	19.77	21.53	11.58	22.24	11.89	22.81	18.74	17.72	15.27	18.91	18.92	18.44	10.77	21.61	**16.24**	16.12	21.07
FTZmT5	21.83	7.98	19.27	**24.68**	10.80	11.92	8.94	23.32	16.82	14.99	12.90	21.01	20.07	15.85	9.14	13.05	11.06	12.66	15.20
Meta-X_{NLG}	**22.83**	**14.02**	**21.54**	24.61	**12.88**	**23.09**	**12.58**	**25.33**	**20.12**	**18.65**	**17.31**	**22.63**	**20.24**	**20.11**	**12.07**	**23.41**	15.45	**17.96**	**22.95**

Wikilingua



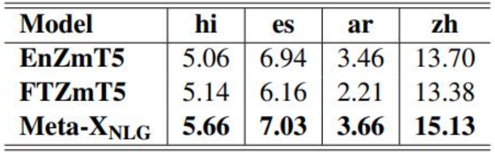
Model	id	fr	ar	pt	it	th	ru	cs	nl	de	ja	zh	hi	tr
EnZmT5	15.34	18.72	**15.70**	17.21	15.05	26.66	14.67	9.42	17.97	13.69	22.32	20.12	18.88	14.45
FTZmT5	13.69	19.37	12.66	17.80	15.54	23.72	11.95	10.20	16.74	12.22	22.81	18.64	17.32	13.84
Meta-X_{NLG}	**16.85**	**20.26**	15.66	**18.36**	**16.03**	**27.71**	**14.89**	**11.76**	**19.09**	**14.11**	**22.83**	**22.45**	**19.60**	**15.23**

MLQA



Model	ar	de	zh	vi	hi	el	ru	ro
EnZmT5	8.55	9.99	23.76	17.29	9.55	8.18	10.98	11.27
FTZmT5	5.82	9.040	22.87	16.47	9.05	6.95	8.87	10.31
Meta-X_{NLG}	**8.63**	**10.52**	**24.89**	**20.92**	**11.90**	**9.01**	**11.41**	**12.24**

Model	fi	ru	id	sw	ko	bn	ta
EnZmT5	7.87	5.52	5.75	4.48	8.59	5.77	3.08
FTZmT5	8.39	7.28	**11.42**	5.51	10.05	7.96	2.022
Meta-X_{NLG}	**9.08**	**7.47**	9.36	**6.42**	**12.67**	**9.17**	**9.76**

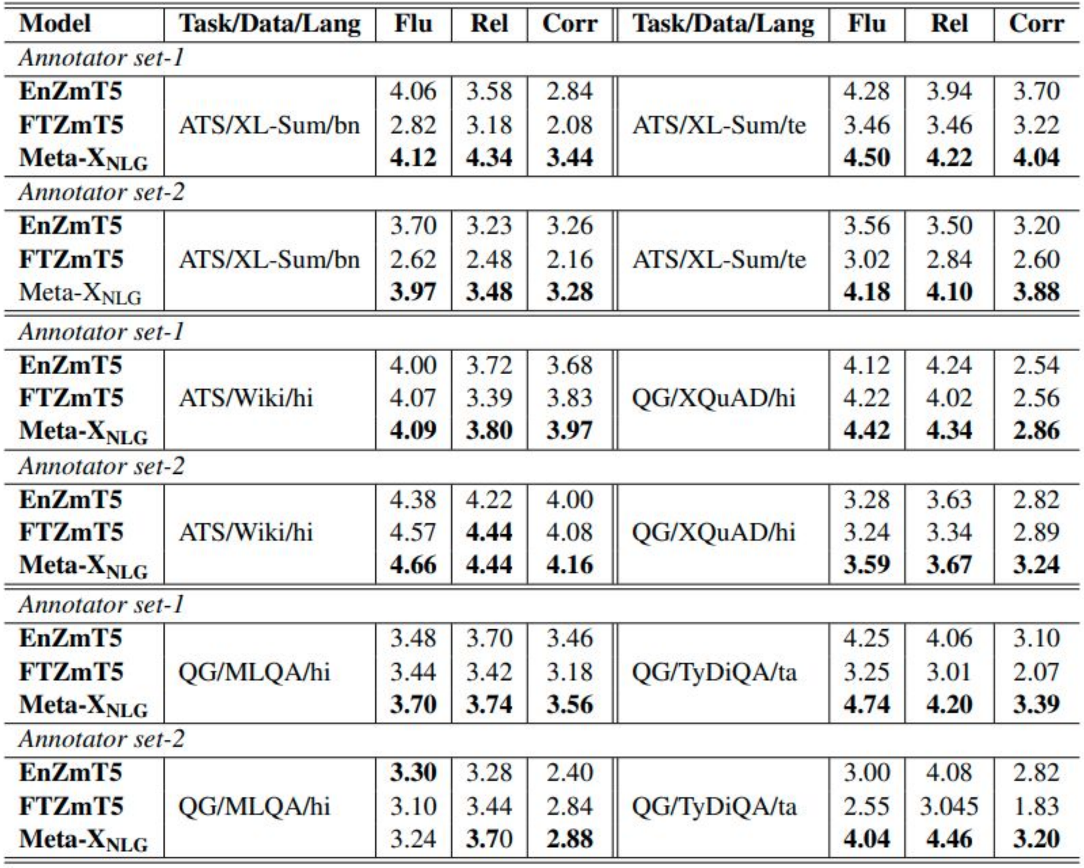


Model	hi	es	ar	zh
EnZmT5	5.06	6.94	3.46	13.70
FTZmT5	5.14	6.16	2.21	13.38
Meta-X_{NLG}	**5.66**	**7.03**	**3.66**	**15.13**

XQuAD

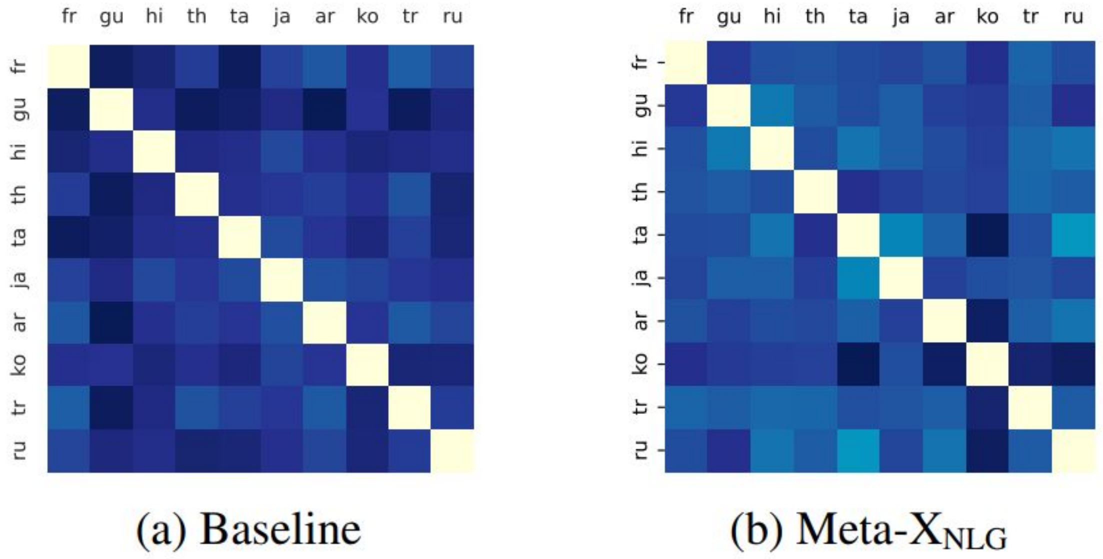
TyDiQA

Results: Human Evaluation Scores



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

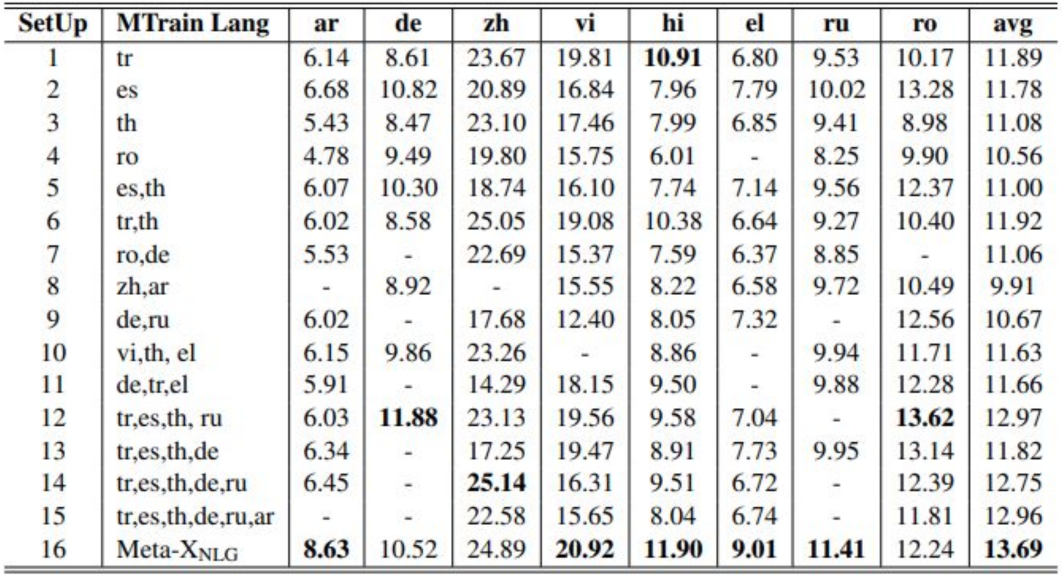
Results Analysis-I: Cross-lingual Transfer



(a) Baseline (b) Meta-X_{NLG}

Cosine distance between language tags obtained from EnZmT5 and Meta-XNLG for 10 languages from XL-Sum dataset. Dark color indicate higher cosine distance.

Results Analysis-II: Effect of Training Languages



SetUp	MTrain Lang	ar	de	zh	vi	hi	el	ru	ro	avg
1	tr	6.14	8.61	23.67	19.81	**10.91**	6.80	9.53	10.17	11.89
2	es	6.68	10.82	20.89	16.84	7.96	7.79	10.02	13.28	11.78
3	th	5.43	8.47	23.10	17.46	7.99	6.85	9.41	8.98	11.08
4	ro	4.78	9.49	19.80	15.75	6.01	-	8.25	9.90	10.56
5	es,th	6.07	10.30	18.74	16.10	7.74	7.14	9.56	12.37	11.00
6	tr,th	6.02	8.58	25.05	19.08	10.38	6.64	9.27	10.40	11.92
7	ro,de	5.53	-	22.69	15.37	7.59	6.37	8.85	-	11.06
8	zh,ar	-	8.92	-	15.55	8.22	6.58	9.72	10.49	9.91
9	de,ru	6.02	-	17.68	12.40	8.05	7.32	-	12.56	10.67
10	vi,th, el	6.15	9.86	23.26	-	8.86	-	9.94	11.71	11.63
11	de,tr,el	5.91	-	14.29	18.15	9.50	-	9.88	12.28	11.66
12	tr,es,th, ru	6.03	**11.88**	23.13	19.56	9.58	7.04	-	**13.62**	12.97
13	tr,es,th,de	6.34	-	17.25	19.47	8.91	7.73	9.95	13.14	11.82
14	tr,es,th,de,ru	6.45	-	**25.14**	16.31	9.51	6.72	-	12.39	12.75
15	tr,es,th,de,ru,ar	-	-	22.58	15.65	8.04	6.74	-	11.81	12.96
16	Meta-X_{NLG}	**8.63**	10.52	24.89	**20.92**	**11.90**	**9.01**	**11.41**	12.24	**13.69**

- Zero-shot results on different training languages combinations of the XQuAD dataset.
- ’-’ indicates the language used in training, so scores are not zero-shot and not included.

Outline

- ❏ Introduction
- ❏ Background: MAML
- ❏ Proposed Approach
- ❏ Experiment Setup
- ❏ Results Analysis
- ❏ Conclusion
- ❏ Other Project: MT for ELRLs

Conclusion

- We propose a novel Meta-XNLG 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 future, we will extend this study to more cross-lingual tasks and languages in the future.

Acknowledgments

- Human Evaluators for human evaluation
- Anonymous reviewers for Constructive feedback
- Microsoft for Travel Grant for attending conference

References

1. Joshi, Pratik, et al. "The state and fate of linguistic diversity and inclusion in the NLP world." *arXiv preprint arXiv:2004.09095* (2020).
2. Wan, Xiaojun, Huiying Li, and Jianguo Xiao. "Cross-language document summarization based on machine-translation quality prediction." *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010
3. Shen, Shi-qi, et al. "Zero-shot cross-lingual neural headline generation." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26.12 (2018): 2319-2327.
4. Duan, Xiangyu, et al. "Zero-shot cross-lingual abstractive sentence summarization through teaching generation and attention." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019.
5. Zhu, Junnan, et al. "NCLS: Neural Cross-Lingual Summarization." *arXiv preprint arXiv:1909.00156* (2019).
6. Kumar, Vishwajeet, et al. "Cross-lingual training for automatic question generation." *arXiv preprint arXiv:1906.02525* (2019).
7. Chi, Zewen, et al. "Cross-Lingual Natural Language Generation via Pre-Training." *arXiv preprint arXiv:1909.10481* (2019).
8. Ayana, S. Shen, Y. Chen, C. Yang, Z. Liu and M. Sun, "Zero-Shot Cross-Lingual Neural Headline Generation," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 12, pp. 2319-2327, Dec. 2018, doi: 10.1109/TASLP.2018.2842432.
9. Lachaux, Marie-Anne, et al. "Unsupervised Translation of Programming Languages." *arXiv preprint arXiv:2006.03511* (2020).
10. Lample, Guillaume, and Alexis Conneau. "Cross-lingual language model pretraining." *arXiv preprint arXiv:1901.07291* (2019).
11. Conneau, Alexis, et al. "Unsupervised cross-lingual representation learning at scale." *arXiv preprint arXiv:1911.02116* (2019).
12. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
13. Lewis, Mike, et al. "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension." *arXiv preprint arXiv:1910.13461* (2019).
14. Liu, Yinhan, et al. "Multilingual denoising pre-training for neural machine translation." *arXiv preprint arXiv:2001.08210* (2020).
15. Artetxe, Mikel, et al. "A Call for More Rigor in Unsupervised Cross-lingual Learning." *arXiv preprint arXiv:2004.14958* (2020).
16. Wu, Shijie, et al. "Emerging cross-lingual structure in pretrained language models." *arXiv preprint arXiv:1911.01464* (2019).
17. Pires, Telmo, Eva Schlinger, and Dan Garrette. "How multilingual is Multilingual BERT?." *arXiv preprint arXiv:1906.01502* (2019).
18. Artetxe, Mikel, Sebastian Ruder, and Dani Yogatama. "On the cross-lingual transferability of monolingual representations." *arXiv preprint arXiv:1910.11856* (2019).

NATURAL LANGUAGE
& INFORMATION
PROCESSING LAB

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

ACL 2022
22ND – 27TH MAY | 60TH MEETING | DUBLIN

Thank You

NLIP Lab

Homepage

Contact us:

✉ cs18resch11003@iith.ac.in
✉ nlip@cse.iith.ac.in
🌐 https://nlip-lab.github.io/
🌐 https://kaushal0494.github.io/