NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

CIKM 2023

GranCATs: Granularity-Specific Contrastive Adapters

Meizhen Liu, Jiakai He, Xu Guo, Jianye Chen,
Siu Cheung Hui, Fengyu Zhou

23th Oct, 2023

Motivation

- MLLMs
  - Using **transformer-based architectures** and been pre-trained on hundreds of languages;
  - Capturing the cross-lingual alignment and **facilitate knowledge transfer**, achieving performance improvements;
  - Cross-lingual knowledge transferring **is not universally effective** on all languages;

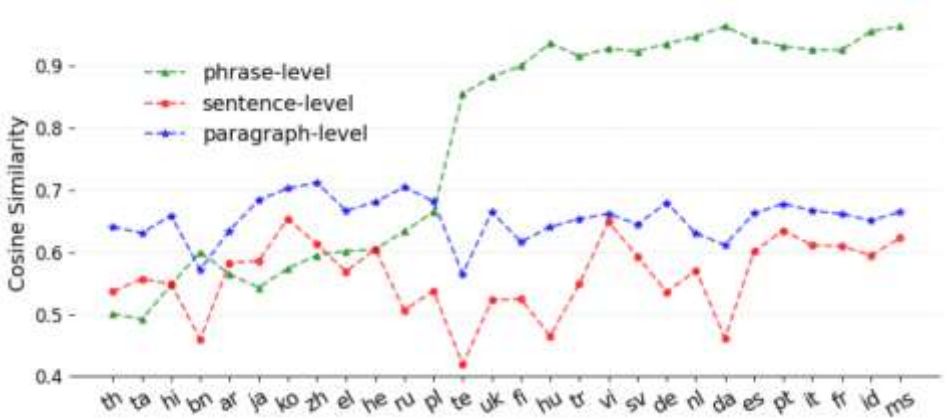
To enhance cross-lingual representations in MLLMs

- Related Works

Full model pre-training on large-scale parallel corpus	→	Computation-consuming and leading to catastrophic forgetting
Employing the pre-training task of MLM, TLM …	→	Local information but not global information
Either entity or sentence level representations	→	Research gap on various text granularities

Preliminary Experiments

- Whether the level of text granularity play an important role in the efficiency of cross-lingual alignments ?
  - 1000 parallel entities;
  - English-center alignments;
  - Covering 29 languages;
  - Cosine similarities;
  - Phrase, sentence and paragraph;

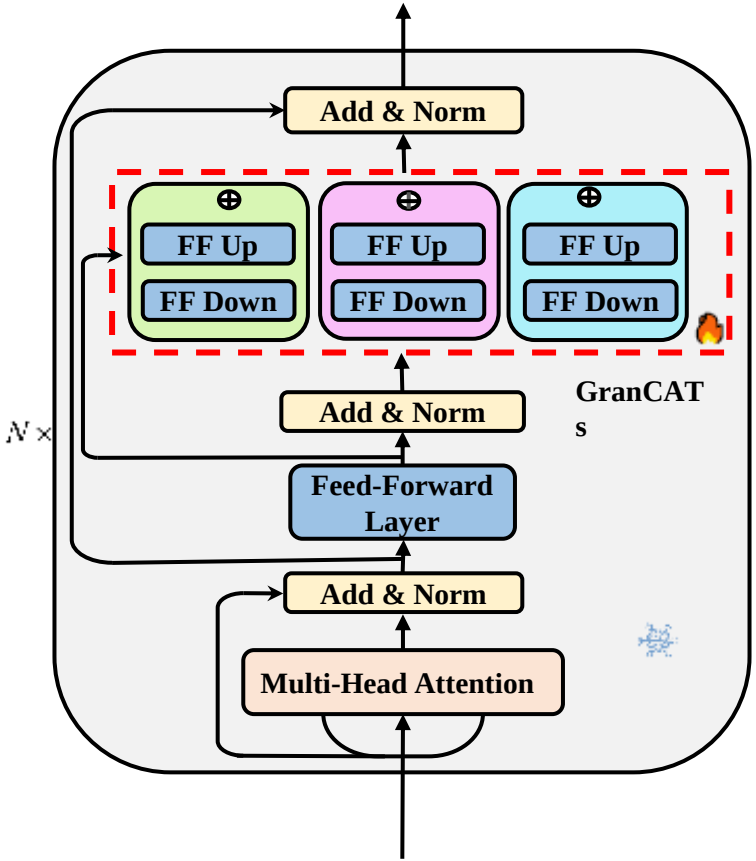


Motivation: Propose a **parameter-efficient method** to enhance the **cross-lingual alignments** across **a range of rext granularities**;

Model Construction Design

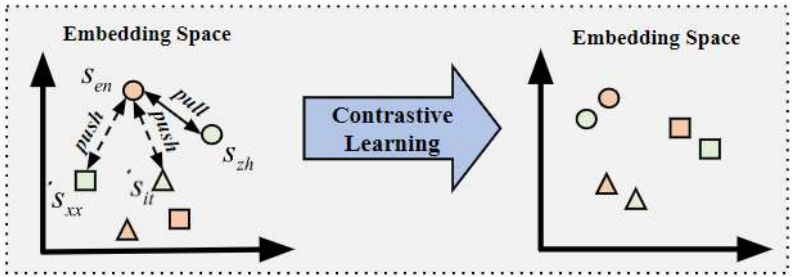
Model Construction

- GranCATs:



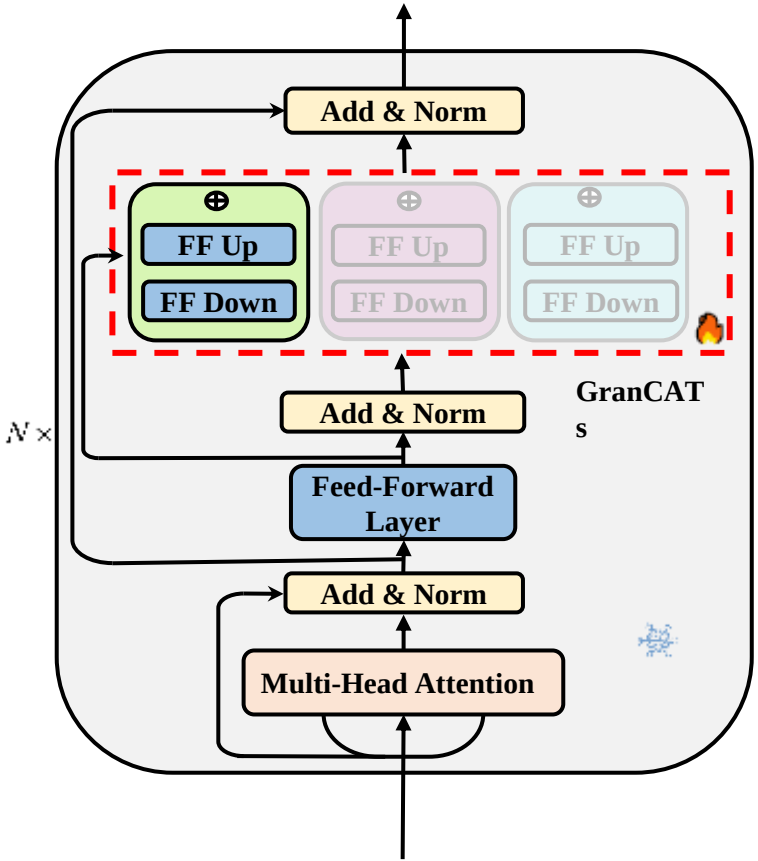
- Contrastive loss with cosine similarity:
  - Focusing on capture the **global information** rather than local semantic alignments; (Park et al,. 2023)
  - CL tents to **pull** the similar samples **close** and **push** distinct samples **away**;

$$\text{InfoNCE}(\boldsymbol{h}_i, \boldsymbol{h}_j) = -\mathbb{E}_N \left[\log \frac{\exp(f(\boldsymbol{h}_i, \boldsymbol{h}_j))}{\sum_{k \in N} \exp(f(\boldsymbol{h}_i, \boldsymbol{h}_k))} \right]$$

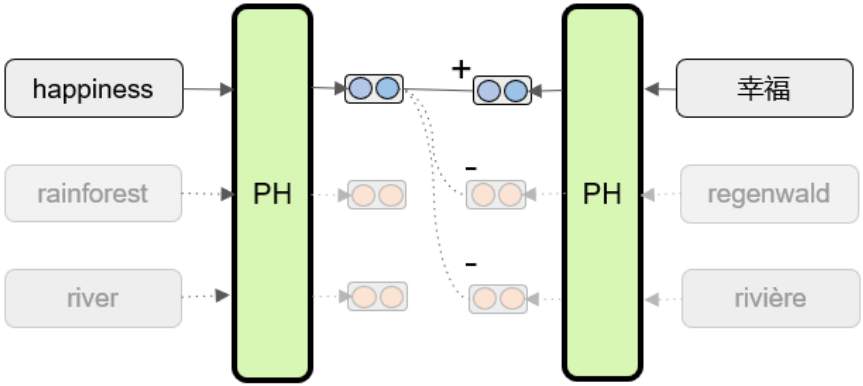


Phrase-level Contrastive Adapter CATs-PH

- For training CATs-PH:



- Taking aligned entities as phrase-level inputs (**in-batch sampling** for negative samples):

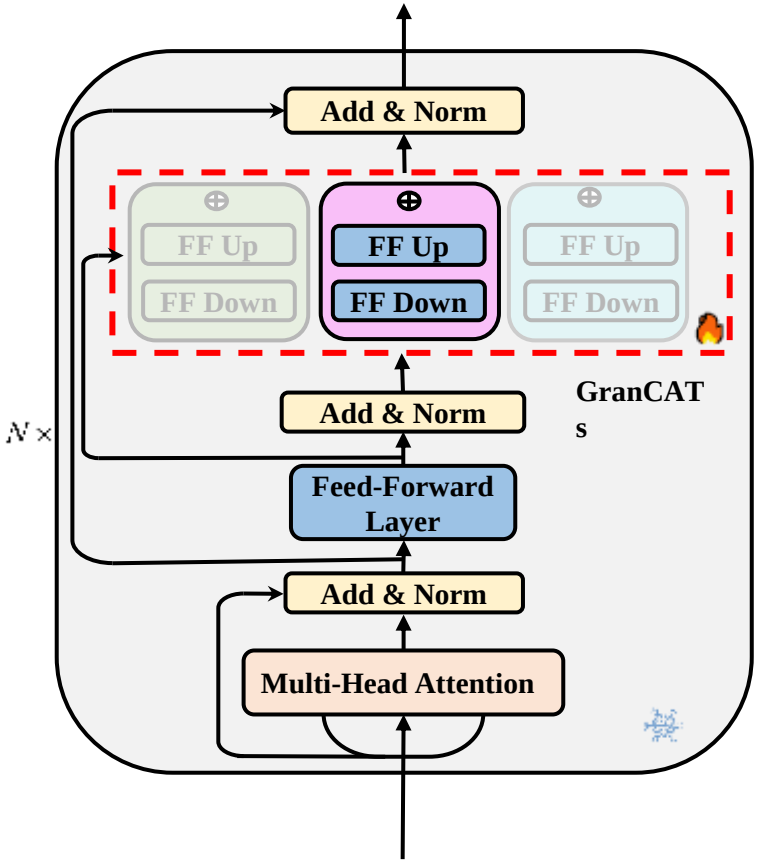


- phrase-level contrastive loss:
  - to grasp the phrase-level cross-lingual alignments;

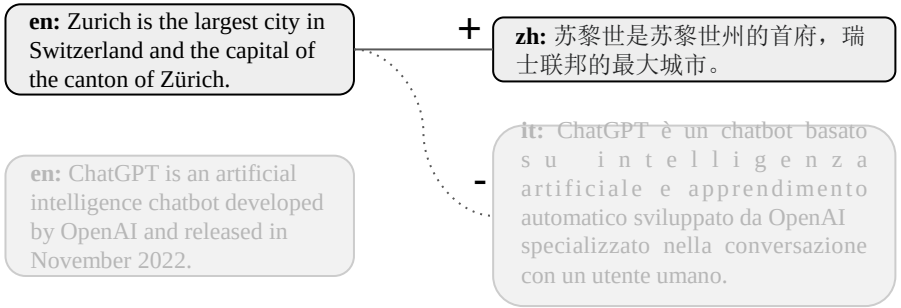
$$\mathcal{L}_{\text{ph}} = \mathop{\mathbb{E}}_{E_{\text{PH}}} \left[\text{InfoNCE}\left(\text{LM}_{\text{PH}}(ph_i^{en}), \text{LM}_{\text{PH}}(ph_i^{k}) \right) \right]$$

Sentence-level Contrastive Adapter CATs-ST

- For training CATs-ST:



- One-sentence description as sentence-level inputs:

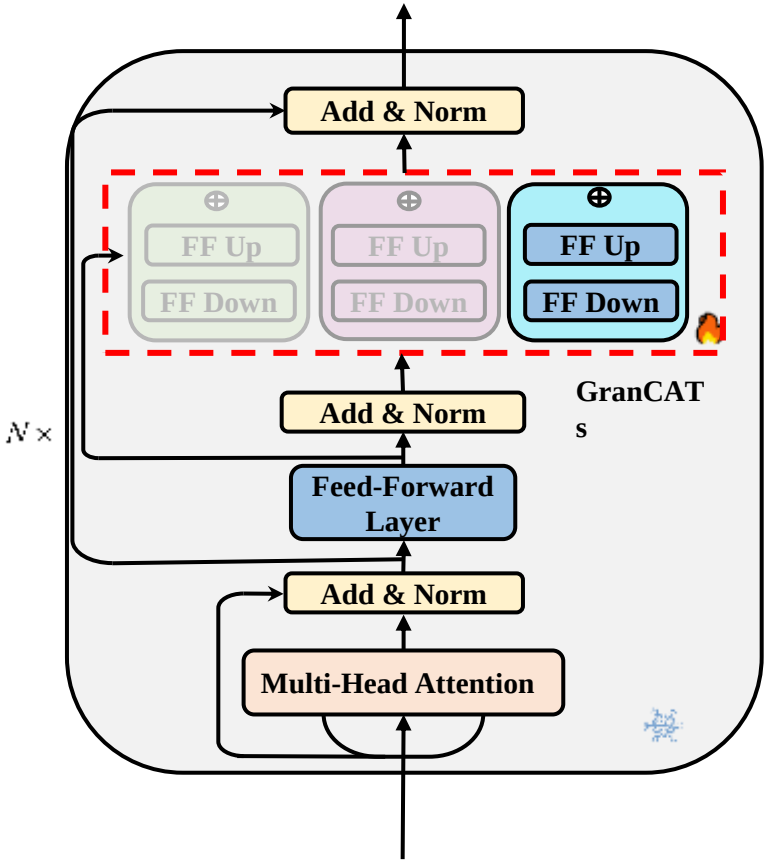


- sentence-level contrastive loss:
  - to calculate the mean of token representations;
  - to use the CL loss;

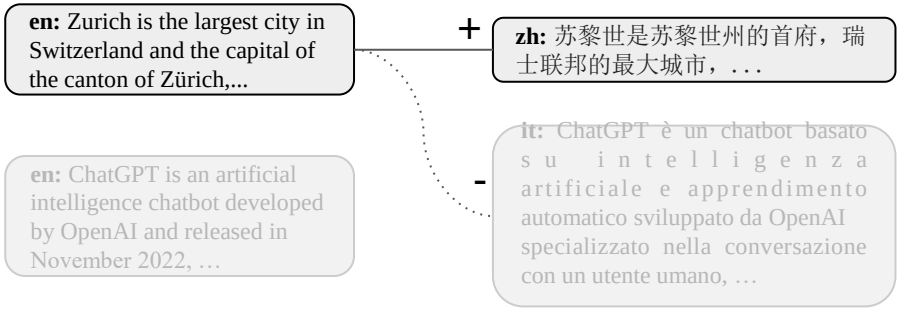
$$\mathcal{L}_{\text{st}} = \mathop{\mathbb{E}}_{E_{\text{ST}}} \left[\text{InfoNCE}\left(\text{LM}_{\text{ST}}(st_i^{en}), \text{LM}_{\text{ST}}(st_i^{k}) \right) \right]$$

Paragraph-level Contrastive Adapter CATs-PG

- Training CATs-PG in the same way:



- The whole description about entities as paragraph-level inputs:



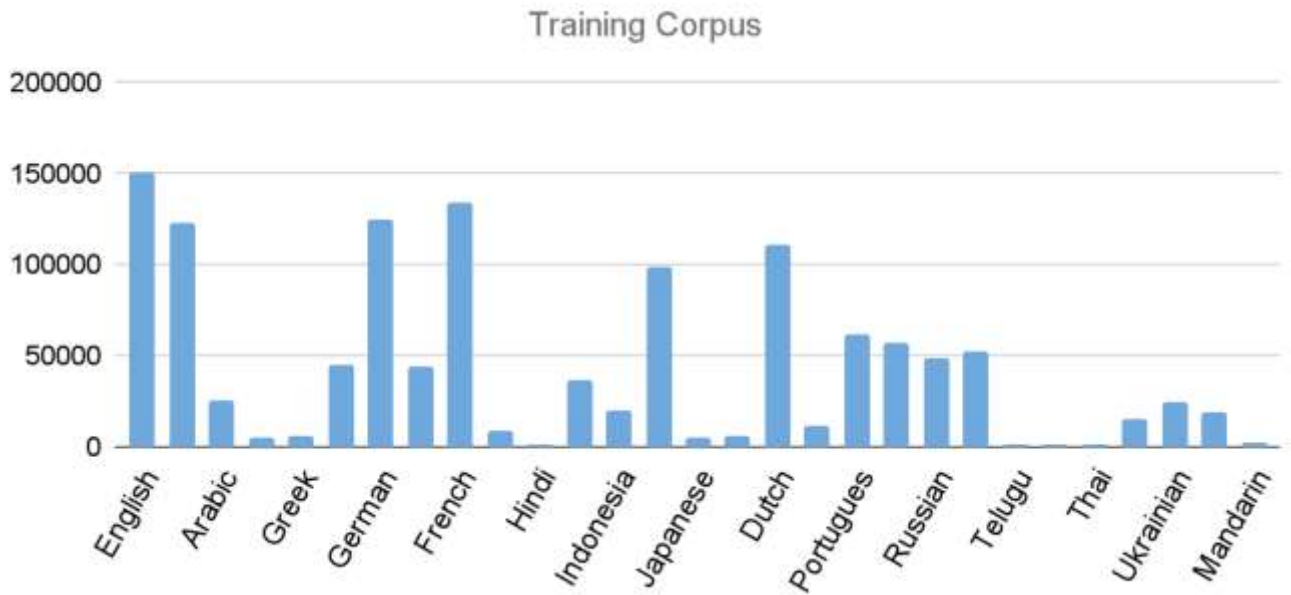
- paragraph-level contrastive loss:

$$\mathcal{L}_{\text{pg}} = \mathop{\mathbb{E}}_{E_{\text{PG}}} \left[\text{InfoNCE}\left(\text{LM}_{\text{PG}}(pg_i^{en}), \text{LM}_{\text{PG}}(pg_i^{k}) \right) \right]$$

The objective is to learn the universal patterns, forcing MLLMs to transfer the alignment knowledge across languages.

Training Corpus Collection

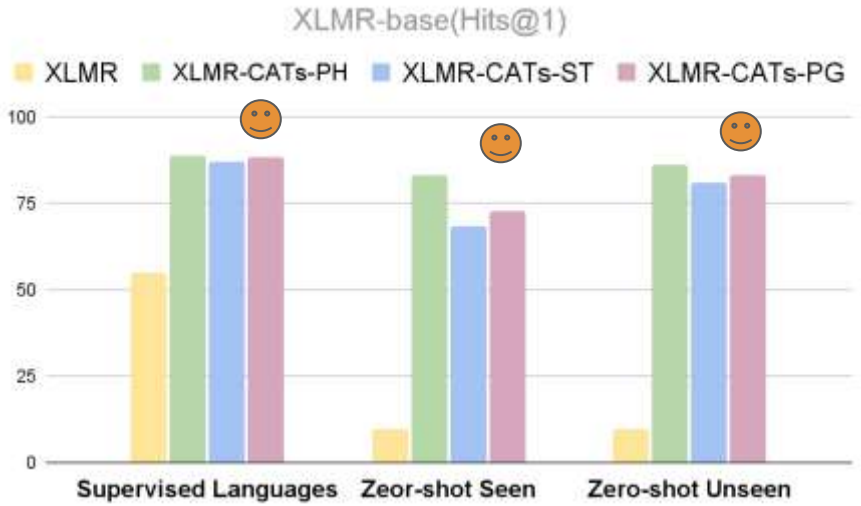
- Collect a large entity-aligned multilingual dataset covering **29** languages from **15** language families, **151K** instances in three kinds of granularities;
- phrase-level & sentence-level & paragraph-level;



Experiment Results on Granularities-Specific Cross-lingual Tasks

Phrase-level Cross-lingual Task—— Entity alignments

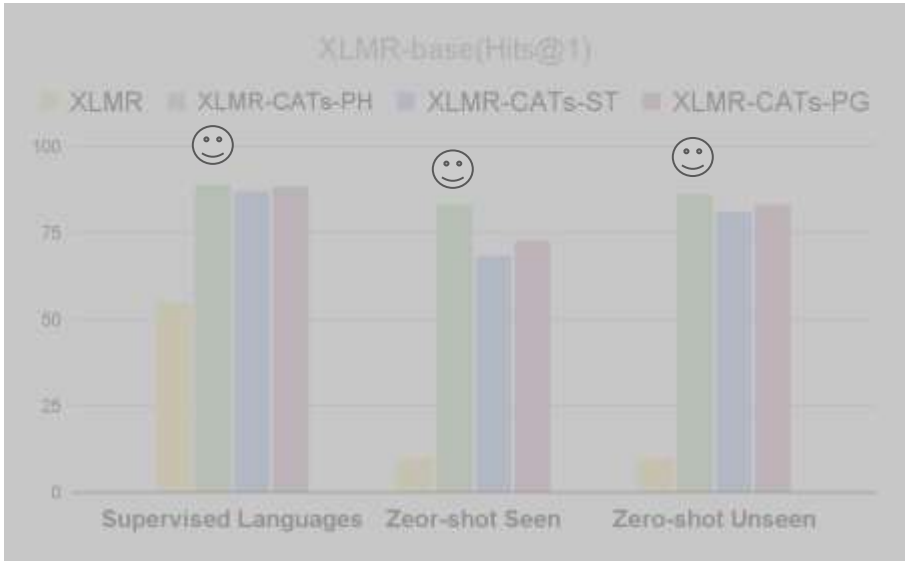
- Entity alignment task (WK3l dataset & Hou et al., 2022)
  - To retrieve the corresponding entity in a target language when given a source language entity;
  - Using on XLMR-Base to make comparison;

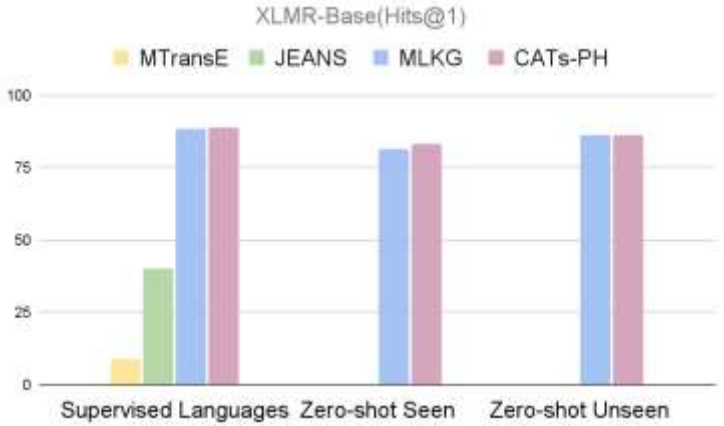


CATs-PH perform best on EA tasks.

Phrase-level Cross-lingual Task—— Entity alignments

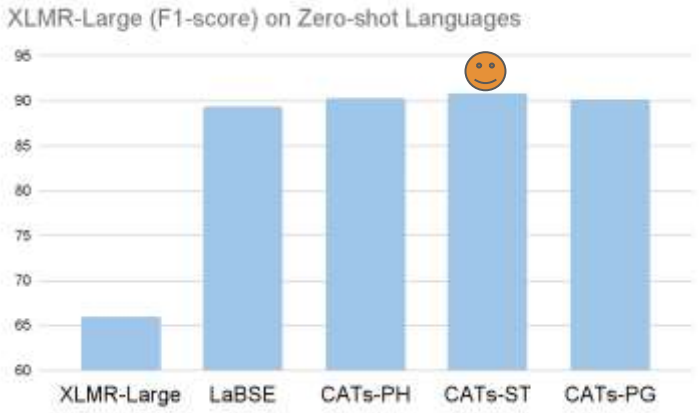
- Entity alignment task (WK3l dataset & Hou et al., 2022)
  - To retrieve the corresponding entity in a target language when given a source language entity;
  - Using XLMR-Base to make comparison;
  - Comparing with the existing MLLMs;





Sentence-level Cross-lingual Tasks

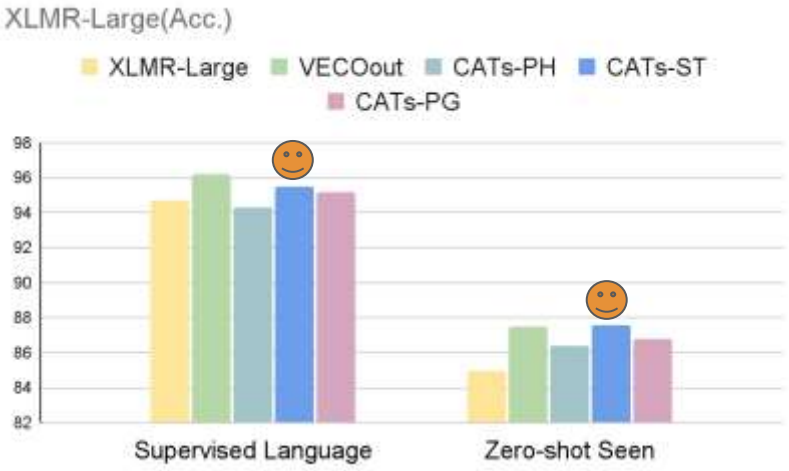
- Cross-lingual Understandings Tasks (XTREME Benchmarks, Hu et al,. 2021)
  - Using the XLMR-Large as the backbone;
  - Sentence retrieval rasks on BUCC dataset;

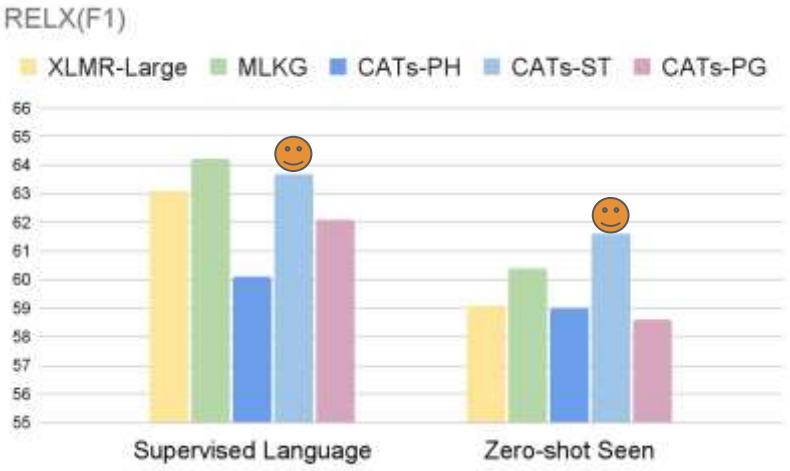


Sentence-level Cross-lingual Tasks

- Cross-lingual Understandings Tasks(XTREME Benchmarks, Hu et al,. 2021)
  - Sentence retrieval rasks on BUCC dataset;
  - Sentence classification task on PAWS-X dataset;
  - Cross-lingual relation extraction task on RELX dataset;

CATs-ST perform best on sentence tasks.

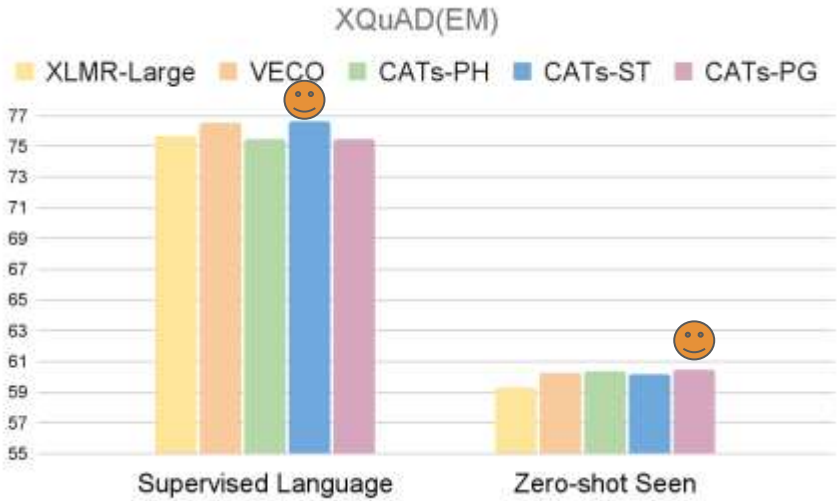


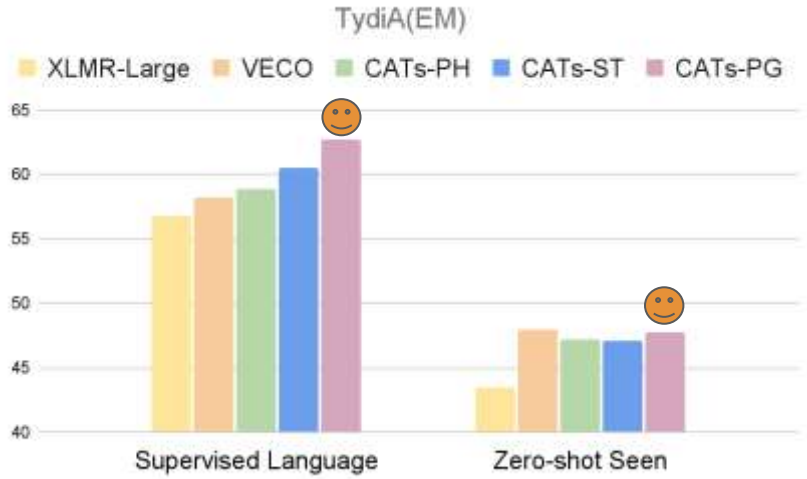


Paragraph-level Cross-lingual Tasks

- Cross-lingual paragraph-level understanding tasks (XTRAME Hu et al., 2021)
  - Question-Answering tasks, XQuAD dataset;
  - Another task on TydiQA dataset;
  - Using XLMR-Large as backbones;

CATs-PG generally perform best.





Analysis Studies

Impact of Pre-training Tasks

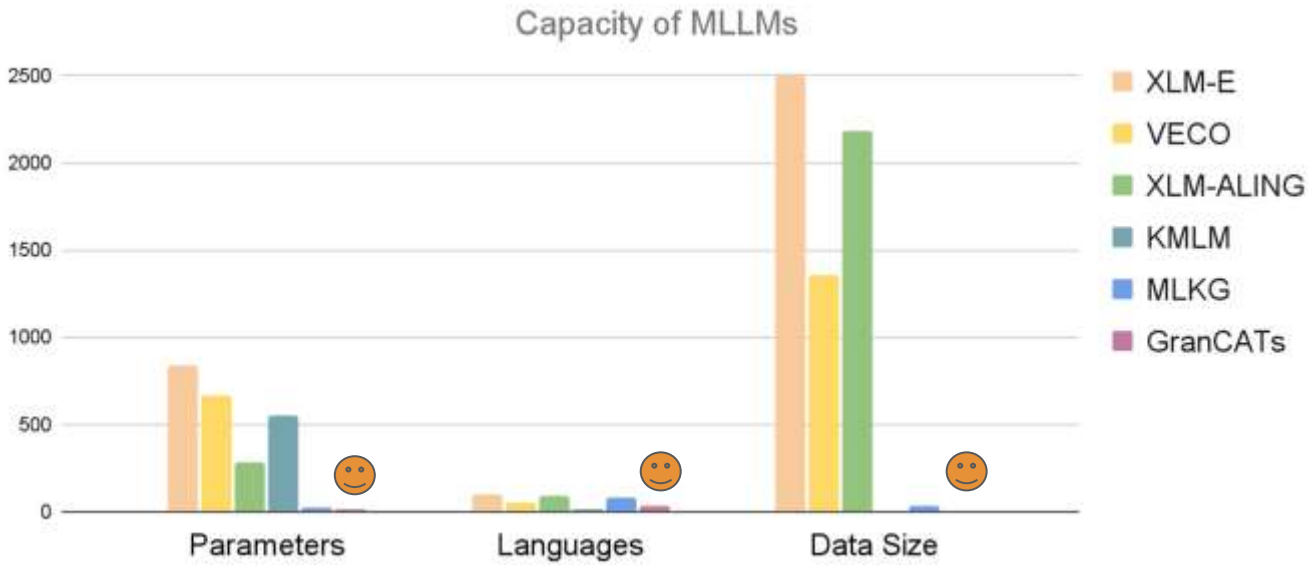
- **NO:** a randomly initialized adapter module without any pre-training;
- **CL:** only with a contrastive loss function;
- **MLM:** only with model language modeling;
- **CL+MLM:** jointly pre-training with both.

Model	WK3l		POS	BUCC	TydiQA	
	Hit@1	MRR	F1	Acc.	F1	EM
No pre-training						
-NO	49.9	50.3	71.3	67.1	57.9	40.6
Pre-training on phrase-level parallel corpus						
-CL	**58.8**	**58.8**	**71.8**	**65.6**	**57.9**	**41.3**
-MLM	44.0	44.3	70.8	60.7	54.9	39.6
-MLM + CL	53.9	54.1	71.0	58.8	56.6	40.8
Pre-training on sentence-level parallel corpus						
-CL	**48.6**	**48.8**	**71.9**	**70.8**	**57.8**	**41.7**
-MLM	39.0	39.3	70.7	64.8	54.6	38.0
-MLM + CL	46.3	46.6	71.3	66.5	55.5	39.5
Pre-training on paragraph-level parallel corpus						
-CL	**44.6**	**44.8**	**71.5**	**69.2**	**58.5**	**42.9**
-MLM	41.4	41.7	71.0	63.2	53.7	37.5
-MLM + CL	44.1	44.4	71.3	65.3	54.7	39.3

CL can effectively capture the cross-lingual information and achieve knowledge transferring.

Comparision of Pre-training Costs

- Pre-training parameters, languages and data size;



Our GranCATs system stands out—it requires fewer parameters and less data. Especially for minority languages with limited parallel.

Takeaways

- Granularities-specific texts have impacts on cross-lingual alignments;
- Enhancement with adapters and contrastive learning works good;

Thanks all the coauthors! This work was completed at ETH Zurich and NTU Singapore and was supported by the China Scholarship Council.

All codes, training corpus and adapter modules will be released to github!

Thanks for your attention!