mPMR: A Multilingual Pre-trained Machine Reader at Scale

Weiwen Xu[1,2], Xin Li[2], Wai Lam[1], Lidong Bing[2]

[1]The Chinese University of Hong Kong

[2]DAMO Academy, Alibaba Group

Motivation

Cross-lingual Language Understanding (XLU)

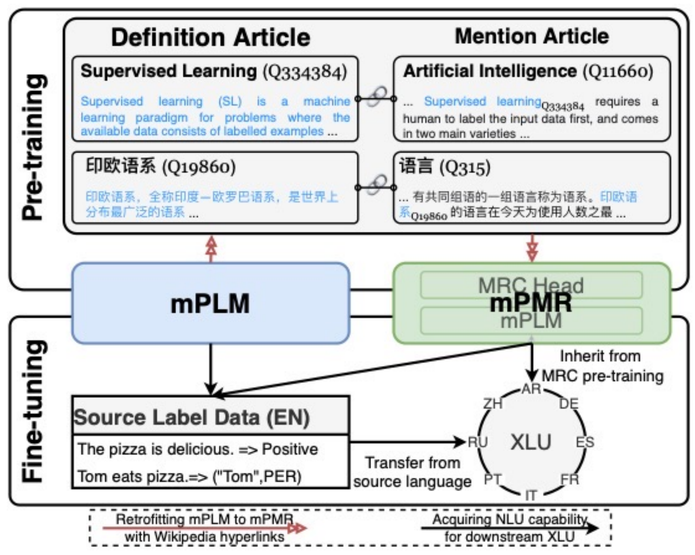
- Fine-tuned on source language data only.
- Perform NLU on multiple target languages.

Multilingual Pre-trained Language Model (mPLM)

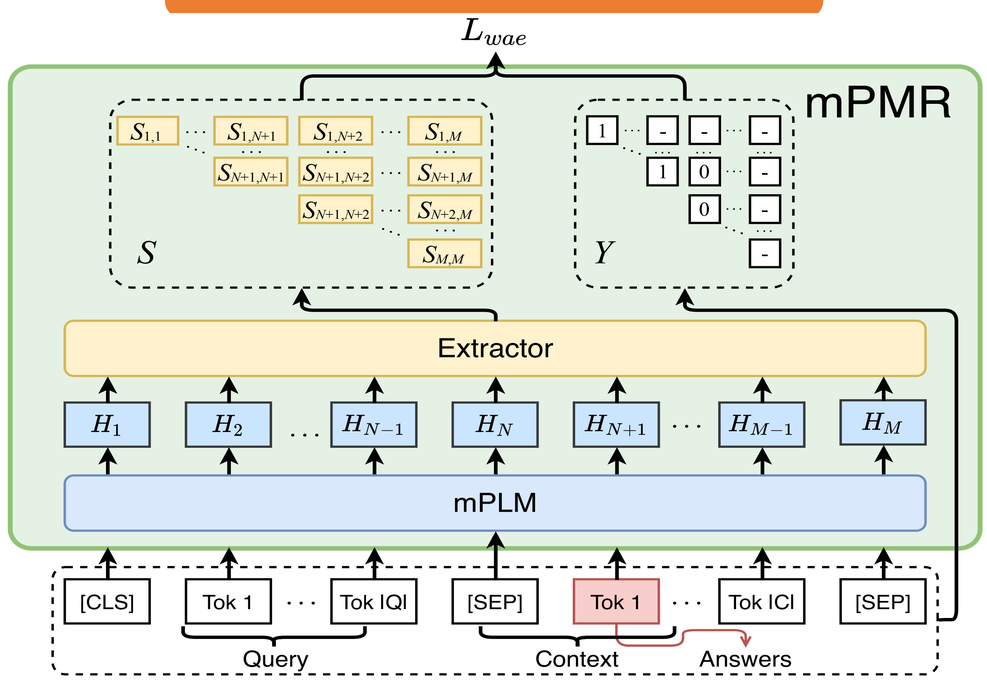
- SRC Language → TGT Language (implicit language alignments)

Multilingual Pre-trained Machine Reader (mPMR)

- SRC Language → TGT Language (implicit language alignments)
- MRC Pre-train → MRC Fine-tune (unified task format)



Model



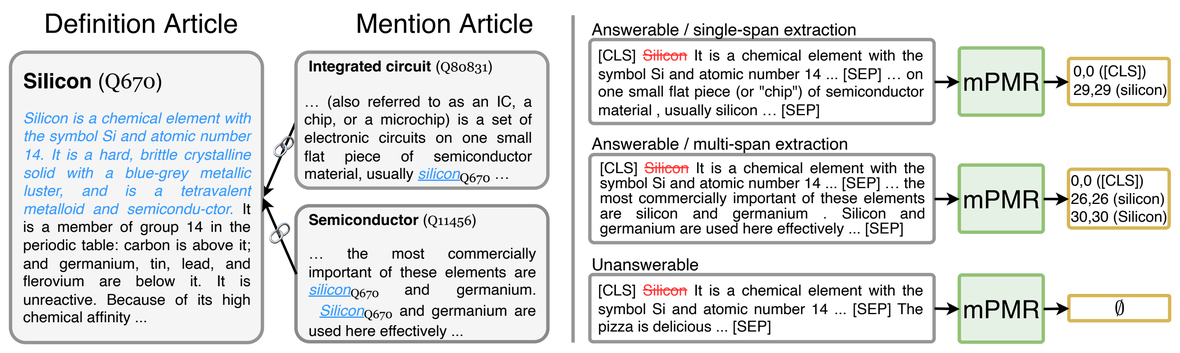
$$X = [[\text{CLS}], Q, [\text{SEP}], C, [\text{SEP}]]$$

$$H = \text{mPLM}(X)$$

$$S = \text{sigmoid}(\boldsymbol{FFN}(H)^T H)$$

$$L_{wae} = \mathbf{CE}(S_{1,1}, Y_{1,1}) + \sum_{N<i\leq j\leq M} \mathbf{CE}(S_{i,j}, Y_{i,j})$$

Data



MRC Triple

- Q: Beginning text from Definition Article.
- C: Surrounding text of the anchor in the Mention Article.
- A: anchor text

Scale to Multiple Languages

- Unified Q/C Construction: Avoid language-specific sentence segmenters and prevent information leakage.
- Stochastic Answer Position: Answer can be presented in any position within the context.

XLU Results

Model	#Params	EQA			NER		ABSA	Sentence Pair		Avg.
		XQuAD	MLQA	TyDiQA	WikiAnn	CoNLL	SemEval16	PAWS-X	XNLI	
Metrics		F1 / EM	F1 / EM	F1 / EM	F1	F1	F1	Acc.	Acc.	
XLM-R	550M	76.6 / 60.8	71.6 / 53.2	65.1 / 45.0	65.4	82.0	66.9‡	86.4	79.2	74.2
mT5	580M	67.0 / 49.0	64.6 / 45.0	57.2 / 41.2	55.7	71.0‡	62.5‡	86.4	75.4	67.5
VECO	550M	77.3 / 61.8	71.7 / 53.2	67.6 / 49.1	65.7	81.3‡	63.0‡	**88.7**	**79.9**	74.4
mLUKE-W	561M	**79.6** / -	72.7 / -	65.2 / 48.5‡	67.7‡	83.0	61.2‡	88.2‡	79.4‡	74.6
Wiki-CL	550M	72.1 / 56.9	70.8 / 50.5	73.2 / 57.3	64.7	-	-	88.4	79.2	-
KMLM	550M	77.3 / 61.7	72.1 / 53.7	67.9 / 50.4	66.7‡	83.2	66.1‡	88.0	79.2	75.1
Our MRC Formulation										
XLM-R$_{base}$	270M	70.8 / 56.9	64.4 / 47.9	50.8 / 38.2	57.9	79.2	60.0	85.0	73.3	67.7
mPMR$_{base}$	270M	74.0 / 59.5	65.3 / 48.7	63.4 / 49.0	66.6	81.7	62.1	86.1	73.6	71.6
XLM-R	550M	77.1 / 61.3	71.5 / 53.9	67.4 / 51.6	63.6	81.4	66.1	86.9	78.6	74.1
mPMR	550M	79.2 / **64.4**	**73.1 / 55.4**	**74.7 / 58.3**	**70.7**	**84.1**	**68.2**	88.0	79.3	**77.2**

Table 1: The results of all XLU tasks. We report the average results of all languages for each dataset. We also compute the overall average score among all datasets in the **Avg.** column. We reproduce the missing results with the ‡ label. Some results of Wiki-CL are left blank because they do not release their model checkpoint.

mPMR Pre-training

Index	Model	#Lang	PAWS-X	XQuAD	WikiAnn	Avg.
#1	XLM-R$_{base}$	0	85.0	70.8	57.9	71.2
#2	#1 + MRC data in English	1	85.2 (0.2↑)	71.0 (0.2↑)	59.5 (1.6↑)	71.9 (0.7↑)
#3	#2 + Stochastic Answer Position	1	85.5 (0.3↑)	73.0 (2.0↑)	60.0 (0.5↑)	72.8 (0.9↑)
#4	#3 + MRC data in more languages	10	85.9 (0.4↑)	73.5 (0.5↑)	64.7 (4.7↑)	74.7 (1.9↑)
#5	#4 + MRC data in even more languages (mPMR$_{base}$)	24	**86.1** (0.2↑)	**74.0** (0.5↑)	**66.6** (1.9↑)	**75.6** (0.9↑)

Table 2: The process of retrofitting XLM-R into mPMR using multilingual MRC data (English→10 languages→24 languages) and our Stochastic Answer Position method. Each row accumulates modifications from all rows above.

mPMR Fine-tuning

Label	Sentence 1	Sentence 2
Entailment	Rami Nieminen (born February 25 , 1966) is a Finnish footballer.	Rami Nieminen (born 25 February 1966) is a Finnish former footballer.
Contradiction	In 1938 he became the Government Anthropologist of the Egyptian-Anglo Sudan and conducted fieldwork with the Nuba.	In 1938 he became the government anthropologist of the anglo-Egyptian Sudan and led fieldwork with the Nuba .
Entailment	Stipsits 出生于科尔新堡，并在维也纳施塔莫斯多夫度过了他的童年。	什蒂普西奇出生于德国科恩堡，在维也纳斯塔莫斯多夫度过了他的童年。
Contradiction	纳舒厄白银骑士团队加入了夏季大学联盟，是本市的现役球队。	Nashua Silver Knights 队是当前夏季联赛的一部分，也是该市的大学体育队。
Entailment	これらの見方は、福音主義的、清教徒的、プロテスタント的な動きが出現するとともに、しばしば表明されてきました。	これらの見解は多くの場合、新教徒、清教徒、福音主義者が出現するなかで示されてきた。
Contradiction	1954 年にスリナムに戻った後、弁護士としてパラマリボに定住した。	1954 年、パラマリボに戻ると、彼はスリナムで弁護士として定住しました。

Table 3: Case study on PAWS-X. mPMR can extract rationales to explain the sequence-pair classification in multiple languages.

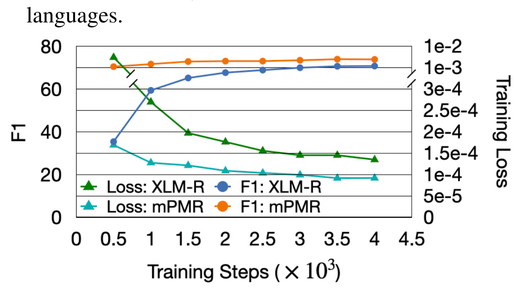


Figure 3: Convergence speed (Test set F1 and the training loss) of mPMR$_{base}$ and XLM-R$_{base}$ on XQuAD.

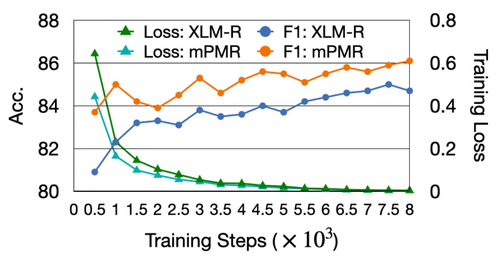


Figure 4: Convergence speed (Test set F1 and the training loss) of mPMR$_{base}$ and XLM-R$_{base}$ on PAWS-X.

Reproducibility

Codes and pre-trained checkpoints are available at https://github.com/DAMO-NLP-SG/PMR