TRELM: Towards Robust and Efficient Pre-training for Knowledge-Enhanced Language Models

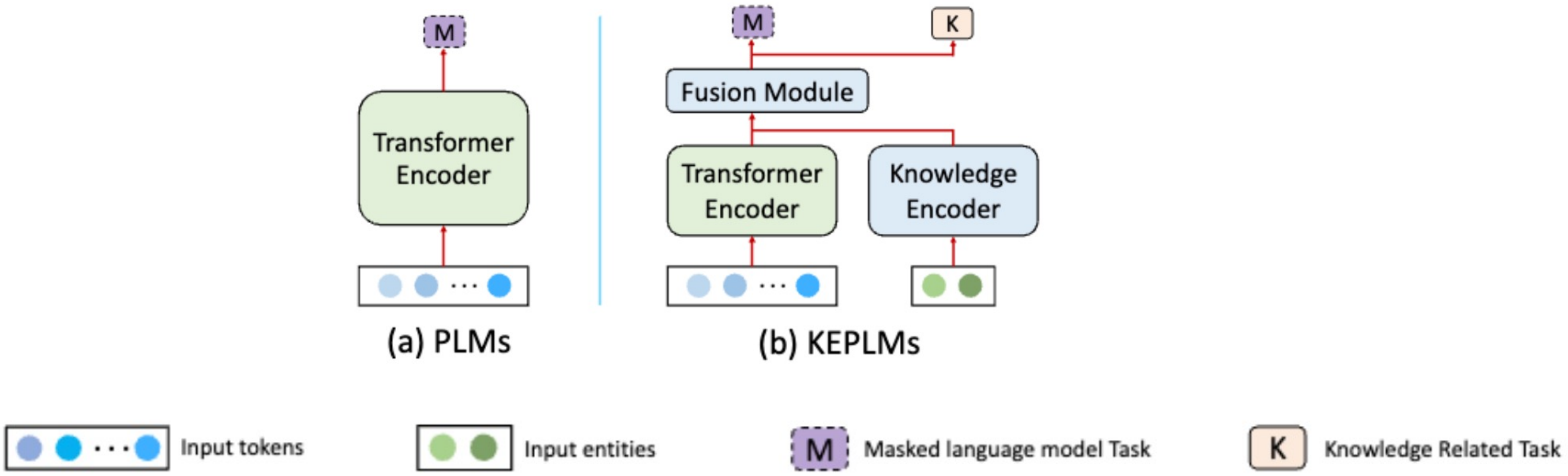
Junbing Yan[1,2], Chengyu Wang[2], Taolin Zhang[2], Xiaofeng He[1], Jun Huang[2], Longtao Huang[2], Hui Xue[2], Wei Zhang[1]

[1] School of Computer Science and Technology, East China Normal University

[2] Alibaba Group

Background & Motivation

- Most of the previous KEPLMs indiscriminately inject knowledge into PLMs, which can introduce noisy knowledge such as redundant or irrelevant information, potentially degrading model performance.
- Some methods modify model backbones with additional knowledge encoders, leading to inflexibility.

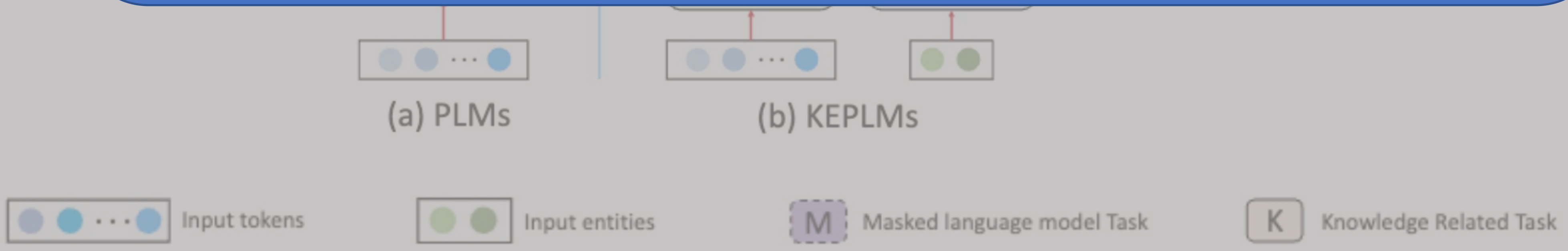


Background & Motivation

- Most of the previous KEPLMs indiscriminately inject knowledge into PLMs, which can introduce noisy knowledge such as redundant or irrelevant information, potentially degra
- Som
  lead

How to avoid injecting entities unrelated to the sentence semantics?

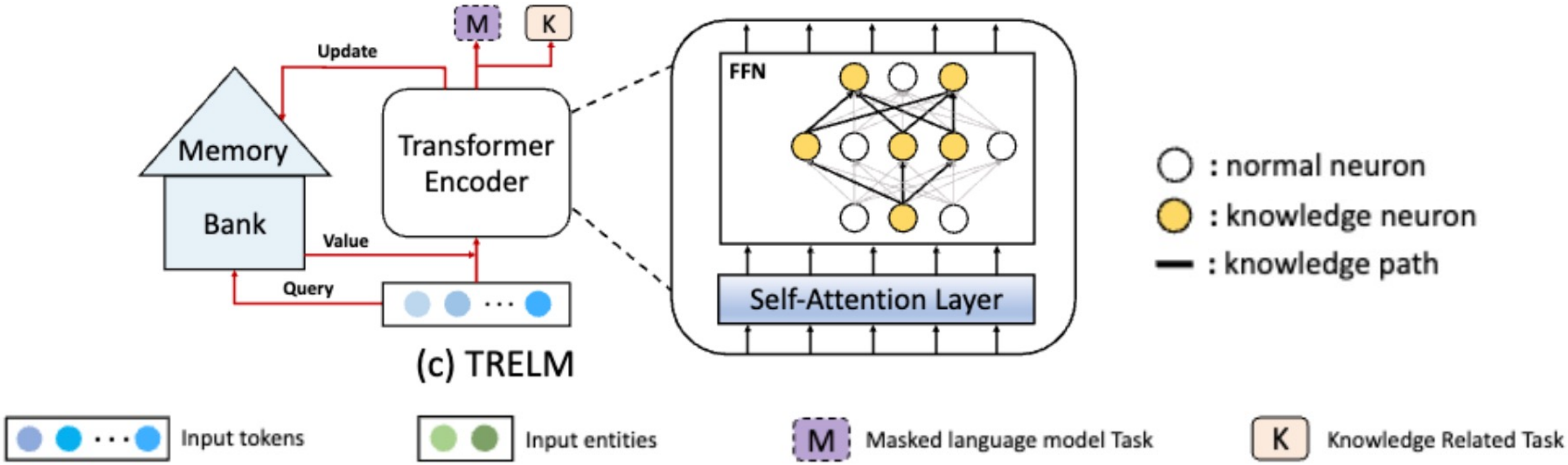
How to inject knowledge without introducing additional parameters?



Background & Motivation

TRELM

- identifying important entities as targets for knowledge injection
- knowledge-augmented memory bank
- selective parameter updates within Transformer blocks



Background & Motivation

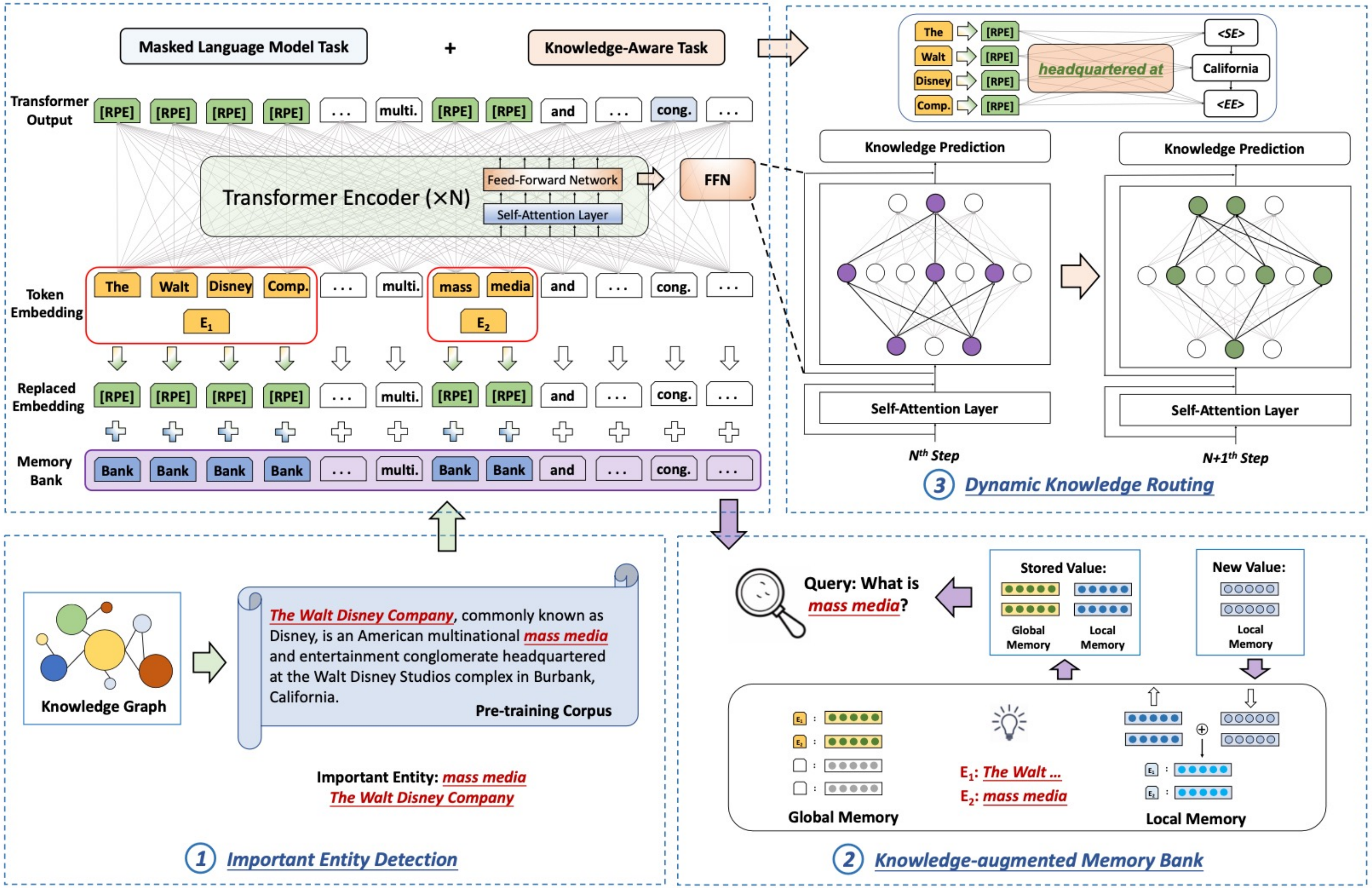
Contribution：

- Knowledge-augmented Memory Bank: Detect important entities in pre-training corpora and construct a knowledge-augmented memory bank, which guides the pre-training process and accelerates convergence.

- Dynamic Knowledge Routing: Propose a novel knowledge routing method that dynamically finds knowledge paths in FFNs and selectively updates model parameters.

- Comprehensive Experiments: Conduct extensive experiments and case studies to show the effectiveness and robustness of TRELM over various NLP tasks.

TRELM Framework



- Important Entity Detection
- Knowledge-augmented Memory Bank
- Dynamic Knowledge Routing

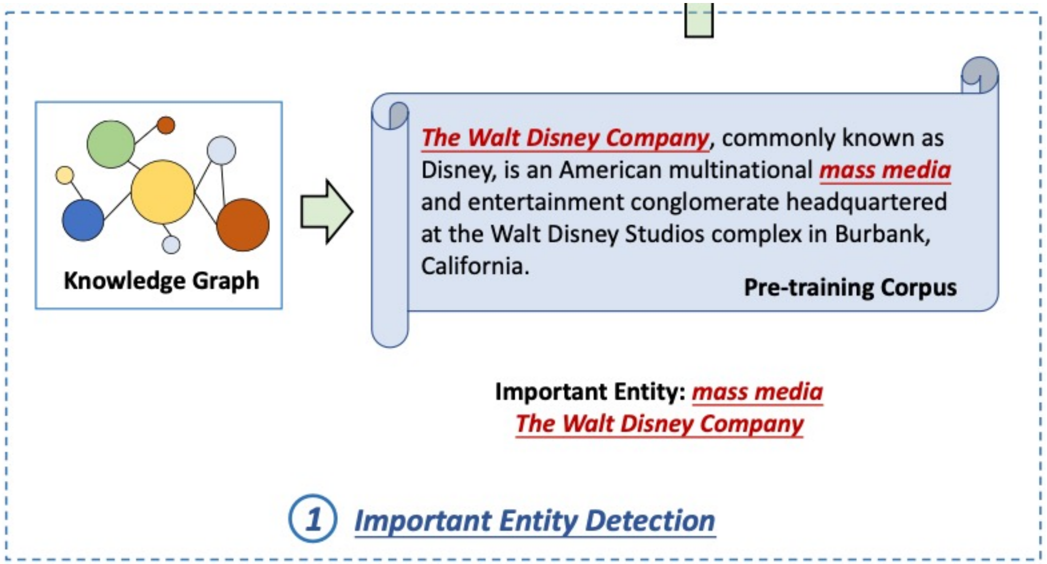
Important Entity Detection

Important Entity Detection：

- Identify which entities in a sentence are more critical for semantic understanding.
- Define the semantic importance score SI(e).

$$SI(e) = \frac{\left\| x_o^T \right\| \cdot \left\| x_{rep} \right\|}{x_o^T \cdot x_{rep}}.$$

- Select entities with higher semantic importance as the target for knowledge injection.



Knowledge-augmented Memory Bank

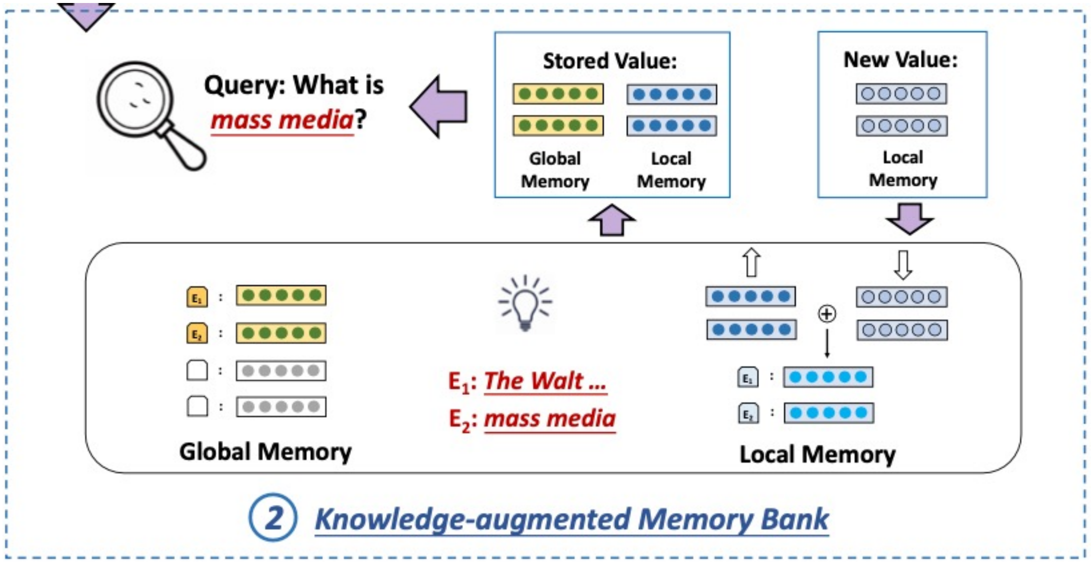
Knowledge-augmented Memory Bank：

- The entities in the corpus follow a long-tail distribution.
- Using a dictionary M to save information around long-tail entities can better optimize the model.

$$\mathcal{M}_{local}^{(e,z)} = \frac{1}{2k+r-l} \sum_{i=l-k}^{r+k} \mathbf{h}_i,$$

- Updating methods of memory bank:

$$\mathcal{M}_{local}^{(e)} \longleftarrow (1-\gamma) \cdot \mathcal{M}_{local}^{(e)} + \gamma \cdot \mathcal{M}_{local}^{(e,z)}$$



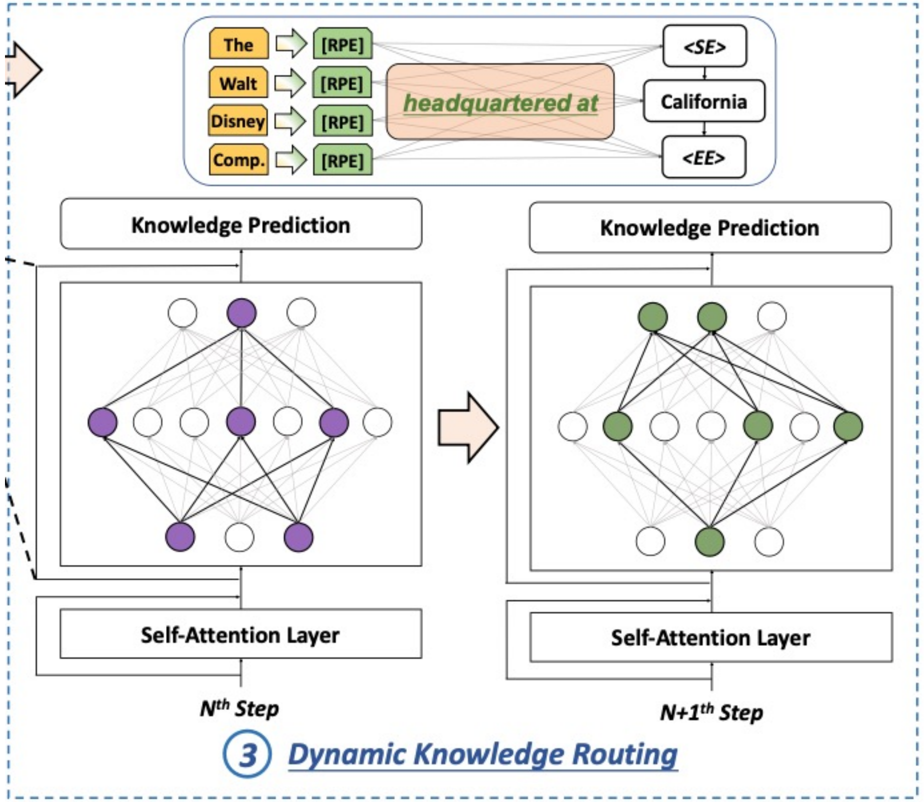
Dynamic Knowledge Routing

Dynamic Knowledge Routing ：

- Some neurons are closely related to knowledge.
- Finding neuron sensitive to knowledge.
- Using knowledge attribution algorithm to calculate the attribution value of neuron.

$$\mathrm{Attr}(v_i^{(l)}) = \overline{v}_i^{(l)} \int_{\alpha=0}^{1} \frac{\partial \, \mathrm{P}_x(\alpha \overline{v}_i^{(l)})}{\partial v_i^{(l)}} d\alpha.$$

- Screening out neurons more sensitive to current knowledge to form a knowledge path.
- Only update this part of the network weight, accelerate model training



Experiment

LAMA Result

Datasets	PLMs	KEPLMs						
	RoBERTa	KEPLER	CoLAKE	KP-PLM	DKPLM	KALM	TRELM	△
T-REx	24.7%	24.6%	28.8%	32.3%	32.0%	29.8%	**33.0%**	+0.7%
UHN-T-REx	17.0%	17.1%	20.4%	22.5%	22.9%	22.6%	**23.3%**	+0.4%
Google-RE	5.3%	7.3%	9.5%	11.0%	10.8%	10.2%	**11.5%**	+0.5%
UHN-Google-RE	2.2%	4.1%	4.9%	5.6%	5.4%	5.2%	**5.9%**	+0.3%

Datasets	BERT	TRELM$_{BERT}$	△
T-REx	32.8%	**36.8%**	+4.0%
UHN-T-REx	23.1%	**28.0%**	+4.9%
Google-RE	11.5%	**15.5%**	+4.0%
UHN-Google-RE	5.8%	**9.7%**	+3.9%

Experiment

Open Entity

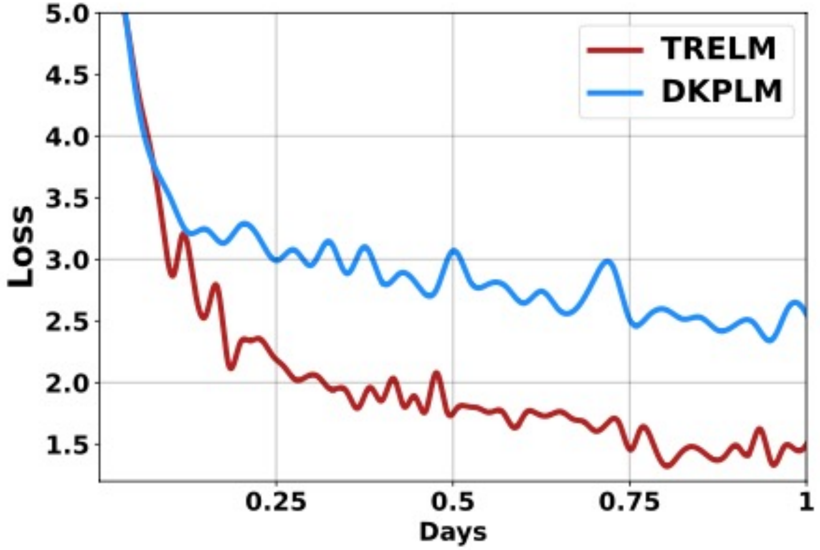
Model	Precision	Recall	F1
BERT	76.4±1.2	71.0±1.4	73.6±1.3
RoBERTa	77.4±1.8	73.6±1.7	75.4±1.8
ERNIE$_{BERT}$	78.4±1.9	72.9±1.7	75.6±1.9
ERNIE$_{RoBERTa}$	80.3±1.5	70.2±1.7	74.9±1.4
KnowBERT	77.9±1.3	71.2±1.5	74.4±1.3
KEPLER$_{WiKi}$	77.8±2.0	74.6±1.9	76.2±1.8
CoLAKE	77.0±1.6	75.7±1.7	76.4±1.5
DKPLM	79.2±1.3	75.9±1.2	77.5±1.2
KP-PLM	**80.8**±1.7	75.1±1.6	77.8±1.7
KALM	78.9±1.5	75.3±1.6	77.1±1.6
TRELM	80.2±1.3	**76.0**±1.4	**78.0**±1.2

TACRED

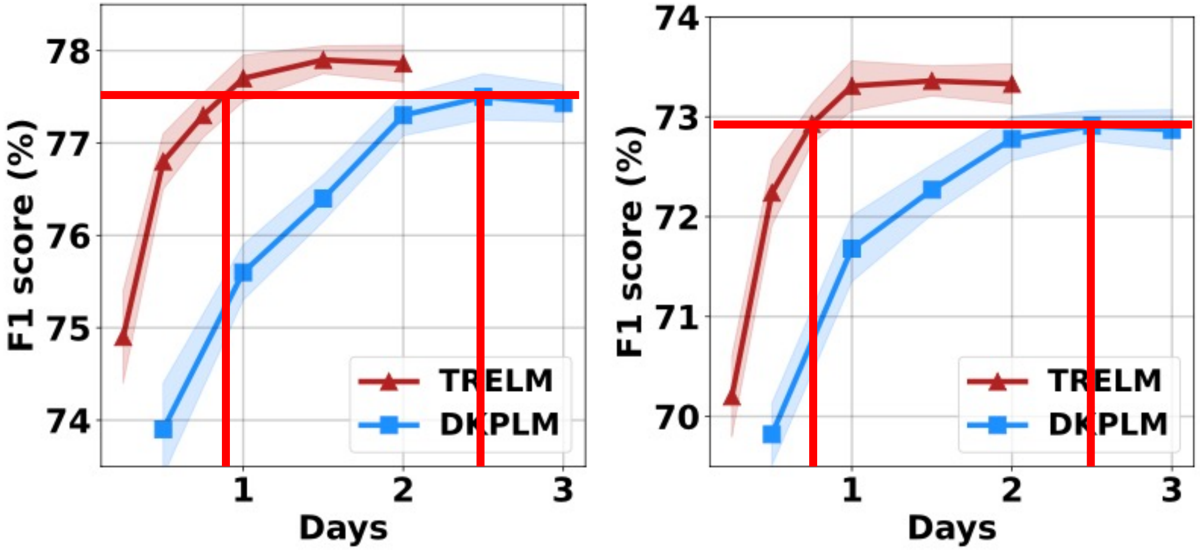
Model	Precision	Recall	F1
BERT	67.23±0.7	64.81±0.6	66.00±0.6
RoBERTa	70.80±0.5	69.60±0.6	70.20±0.5
ERNIE	70.01±0.8	66.14±0.7	68.09±0.7
KnowBERT	71.62±0.7	71.49±0.6	71.53±0.8
DKPLM	72.61±0.5	73.53±0.4	73.07±0.5
KP-PLM	72.60±0.8	73.70±0.7	73.15±0.7
KALM	72.52±0.8	73.38±0.9	72.95±0.8
TRELM	**72.89**±0.5	**73.84**±0.4	**73.36**±0.4

Experiment

Accelerate Performance



Required training time is only 40% of the original SOTA algorithm to achieve the same performance.



Ablation Study

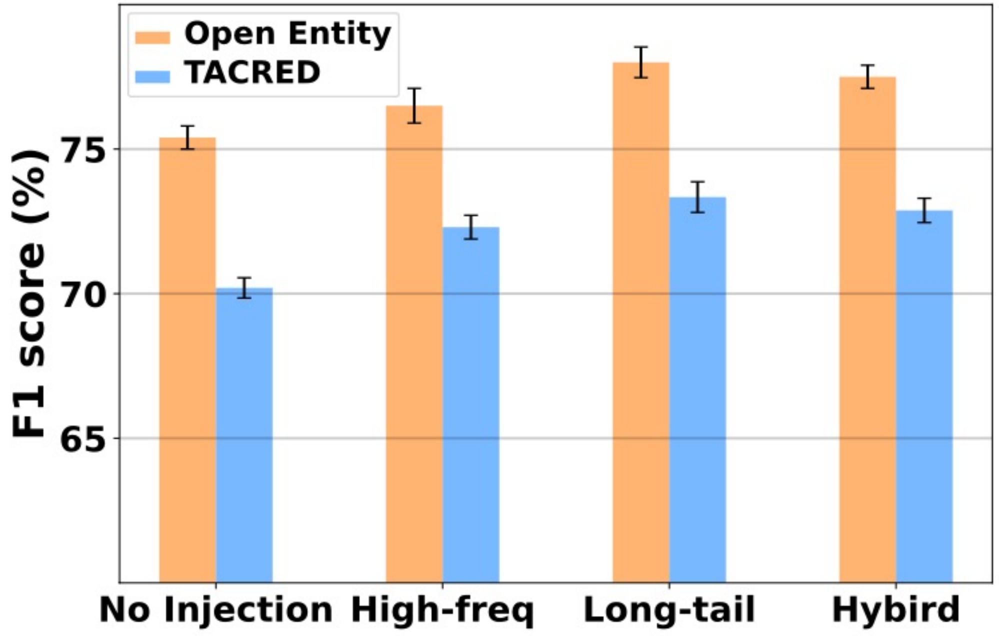
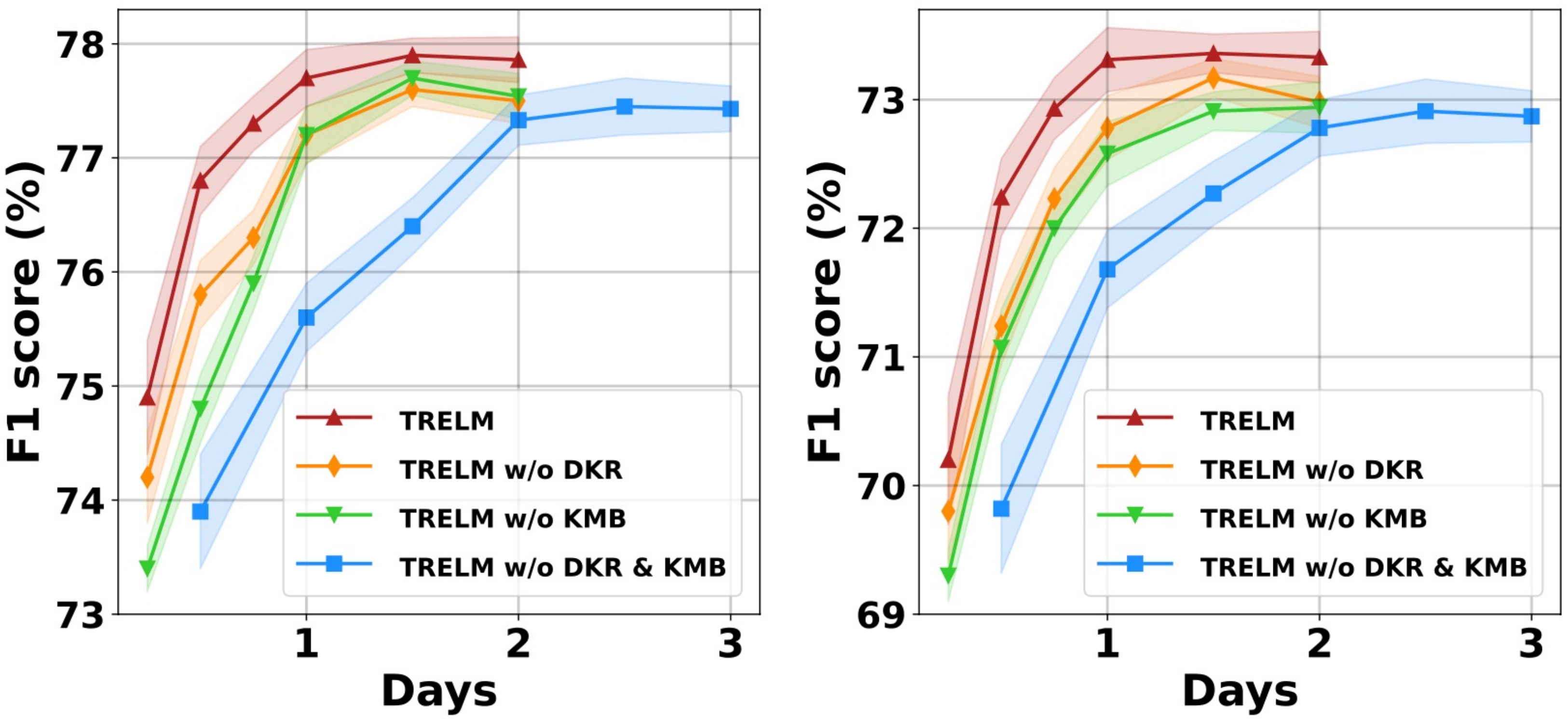


Figure 3: Injection method efficiency over Open Entity and TACRED.

(1) Injecting knowledge into long-tail entities yields better results than limiting it to high-frequency entities, suggesting a greater benefit in enriching representations for entities with sparse occurrences.

(2) Superior performance can be achieved by selectively incorporating knowledge into specific subsets of entities, rather than indiscriminately targeting all available entities.

Ablation Study



(1) Both KMB and DKR enhance the convergence rate of TRELM in the pre-training phase.

(2) KMB exhibits a more pronounced effect on expediting training in the early stages, while DKR's influence becomes increasingly significant over time, ultimately contributing to a greater overall efficiency.

Conclusion

We propose TRELM, a robust and efficient training paradigm for pre-training KEPLMs.

TRELM introduces two innovative mechanisms designed to streamline the integration of knowledge into PLMs without requiring extra parameters:

(1) a knowledge-augmented memory bank that prioritizes knowledge injection for important entities.

(2) a dynamic knowledge routing method that accelerates KEPLMs training and enhances language understanding by updating only the knowledge paths associated with factual knowledge.

華東師範大學 EAST CHINA NORMAL UNIVERSITY

Alibaba

LREC-COLING 2024

Thanks