Building Natural Language Processing Applications with EasyNLP

Chengyu Wang, Minghui Qiu, Jun Huang

Alibaba Group, Hangzhou, China

Main Contents

✓**Knowledge-enhanced Pre-training**

✓Deploying Large Pre-trained Models

- Prompt-based Few-shot Learning
- Knowledge Distillation for Large Pre-trained Models

✓Multi-modal Pre-trained Models

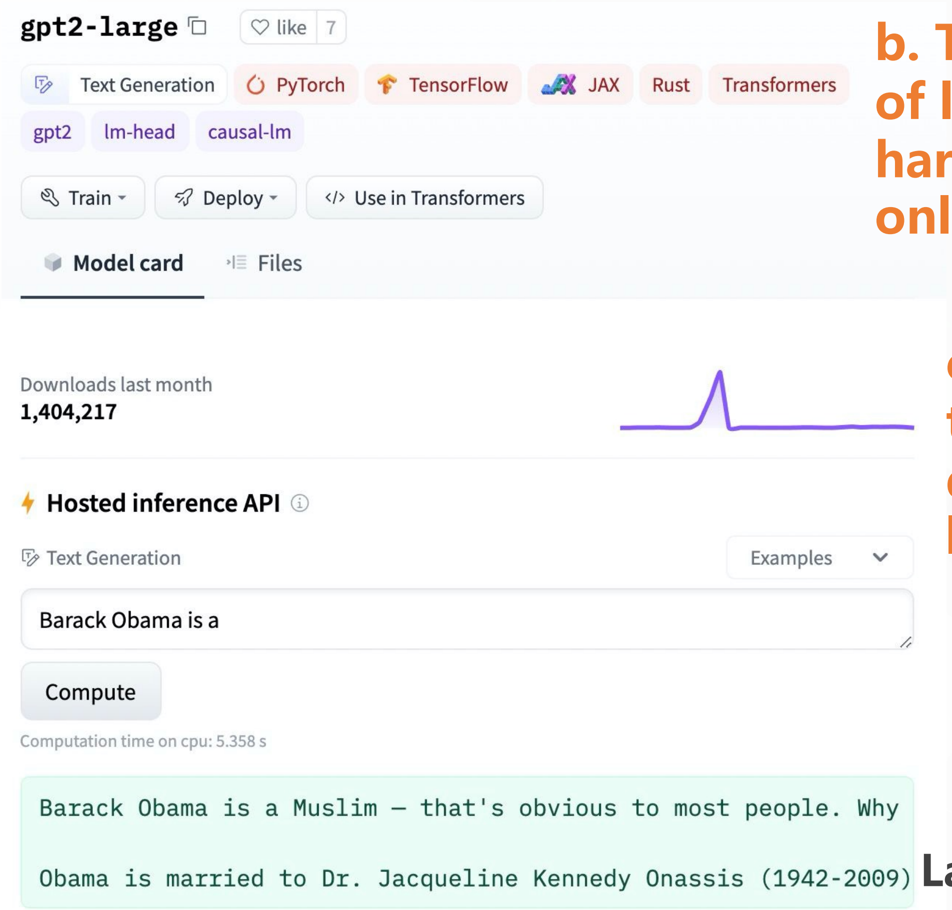
✓Overview of EasyNLP

Development and Challenges for Pre-trained Models

Larger pre-trained models often lead to better performance.

Rank	Name	Model	URL	Score
1	Liam Fedus	SS-MoE		91.0
2	Microsoft Alexander v-team	Turing NLR v5		90.9
3	ERNIE Team - Baidu	ERNIE 3.0		90.6
+ 4	Zirui Wang	T5 + UDG, Single Model (Google Brain)		90.4
+ 5	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.3

Yet, it is not easy to apply large pre-trained models to real-world, industrial applications.



b. The low inference speed of large models make them hard to be deployed online.

c. Large models are easy to overfit, and are difficult to train with little training data.

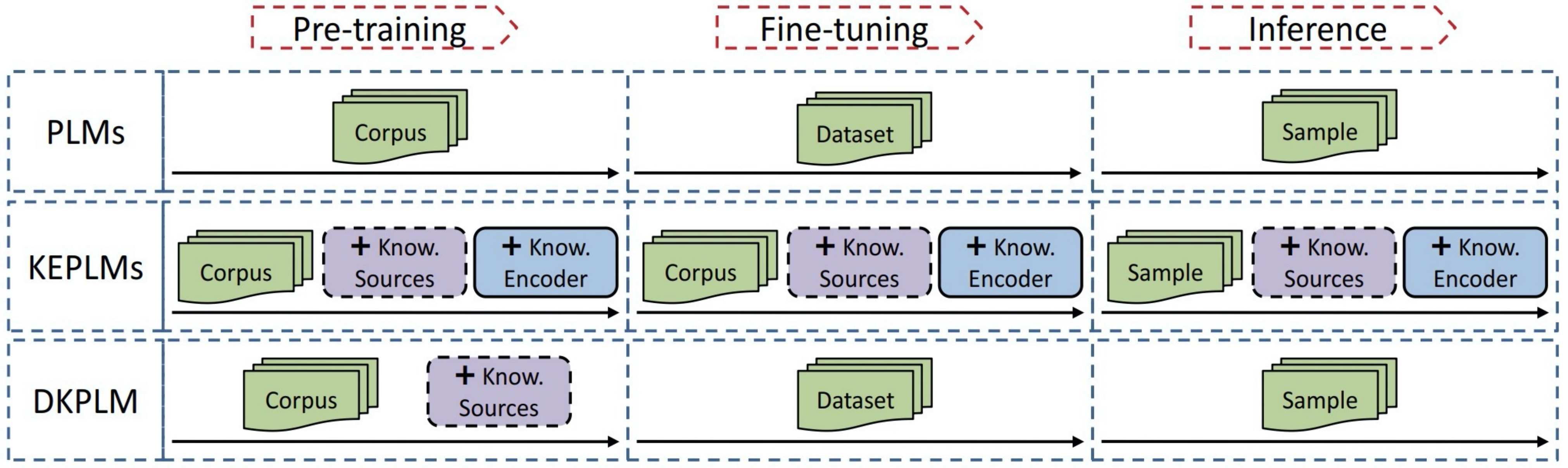
Big Model & Small Labeled Data = OVER FITTING

a. Large models are black boxes, which are prone to anti-common sense errors. The prediction performance in specific domains is also poor.

DKPLM (Decomposable Knowledge-injected PLM)

Main Features of DKPLM

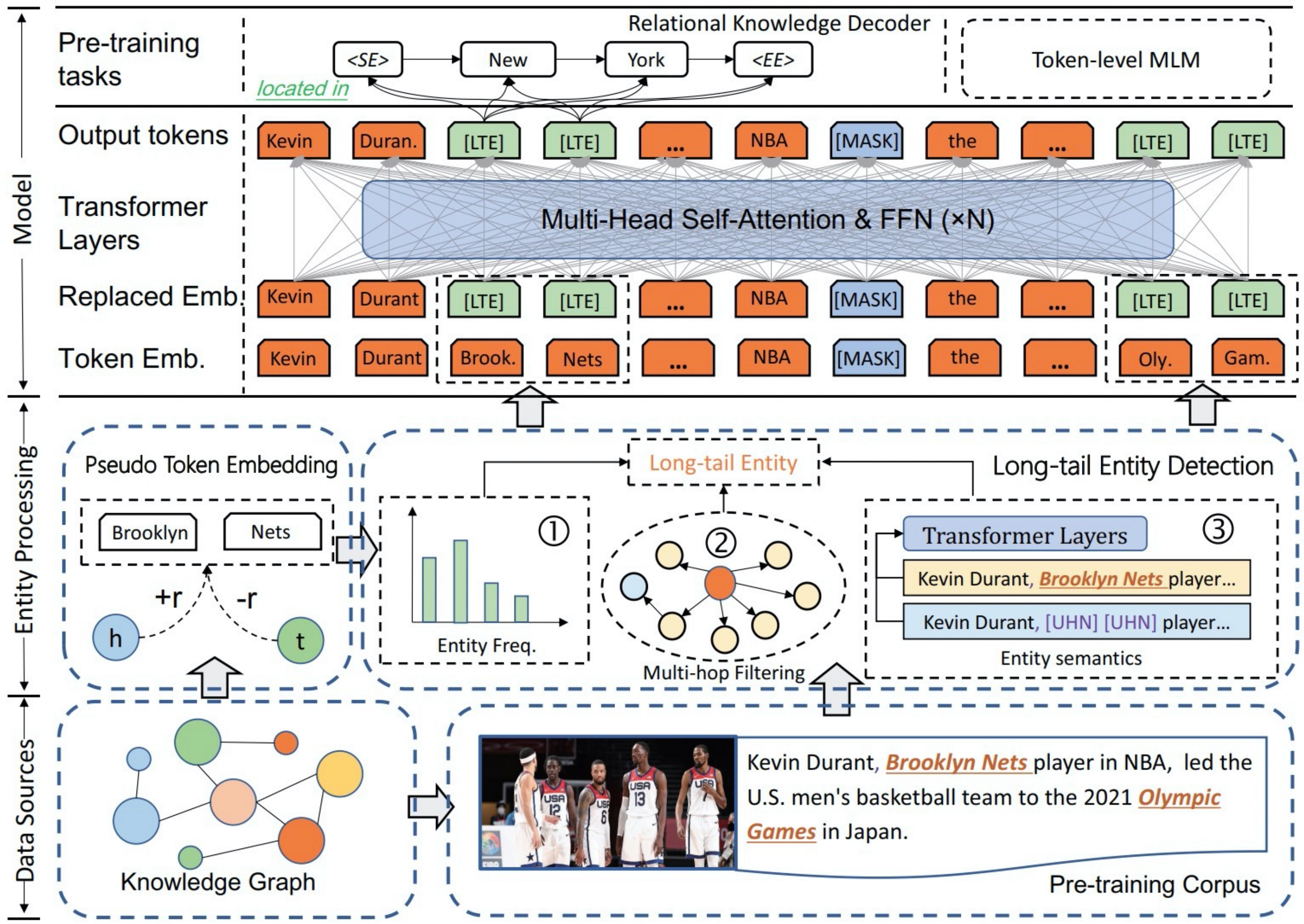
- DKPLM only uses knowledge graphs in pre-training, which is easy to tune and deploy during fine-tuning and inference.
- It effectively protects the knowledge graph data and avoids leakage for cloud service.
- The structure of DKPLM is compatible with BERT and can be directly used by the open-source community.



Taolin Zhang*, Chengyu Wang*, Nan Hu, Minghui Qiu, Chengguang Tang, Xiaofeng He, Jun Huang. DKPLM: Decomposable Knowledge-enhanced Pre-trained Language Model for Natural Language Understanding. AAAI 2022

DKPLM for Knowledge-enhanced Pre-training

Framework of DKPLM



Key Techniques

- **Knowledge injection for long-tail entities**
  - Avoiding learning too much redundant knowledge
- **No additional parameters**
  - Making the backbone fully aligned with BERT
- **Relation-based knowledge decoding**
  - Decoding the injected triple knowledge as one of the pre-training tasks

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{MLM}} + (1 - \lambda_1)\mathcal{L}_{\text{De}}$$

Evaluation Results

Our medical DKPLM

	DKPLM	BERT
CMedQANER (NER)	**84.79**	81.43
CHIP20 (RE)	**77.13**	73.05
CMedMRC (MRC)	**EM=67.18 F1=85.33**	EM=66.15 F1=84.08

Our financial DKPLM

	DKPLM	BERT
FinNER (NER)	**87.81**	77.56
FinSent (Sentence Classification)	**85.75**	83.68
FinMatch (Sentence Matching)	**92.81**	91.99
FinNegReview (Sentence Classification)	**93.81**	92.50

Hugging Face Models

alibaba-pai/**pai-dkplm-medical-base-zh**

alibaba-pai/**pai-dkplm-financial-base-zh**

Hosted inference API

Fill-Mask

Examples

Mask token: [MASK]

感冒需要吃[MASK]

Compute

Computation time on cpu: 0.077 s

药	0.938
吗	0.012
的	0.009
?	0.007
。	0.006

JSON Output

Maximize

Main Contents

✓Knowledge-enhanced Pre-training

✓**Deploying Large Pre-trained Models**

- Prompt-based Few-shot Learning
- Knowledge Distillation for Large Pre-trained Models

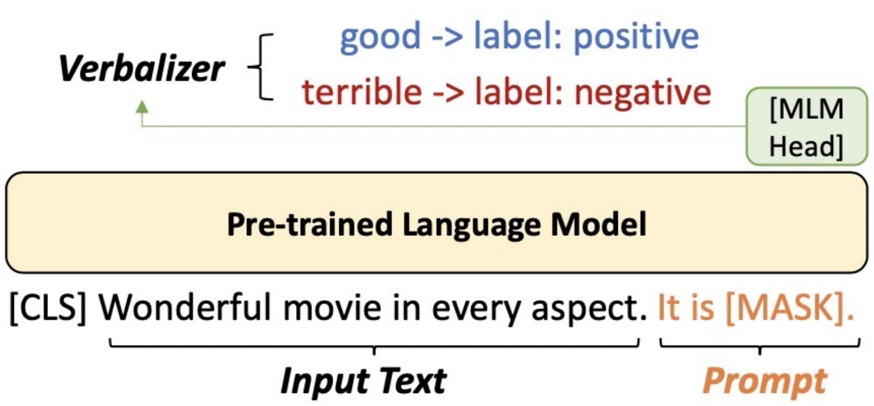
✓Multi-modal Pre-trained Models

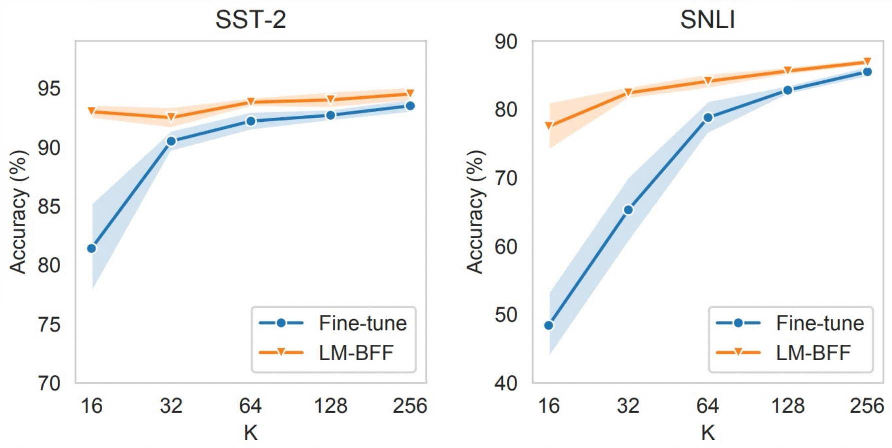
✓Overview of EasyNLP

Why Prompt-based Few-shot Learning?

✓ **Fine-tuning**: requires sufficient labeled training data, hard to obtain in some real-world applications

✓ **Prompt-based Fine-tuning**: a new paradigm for few-shot learning



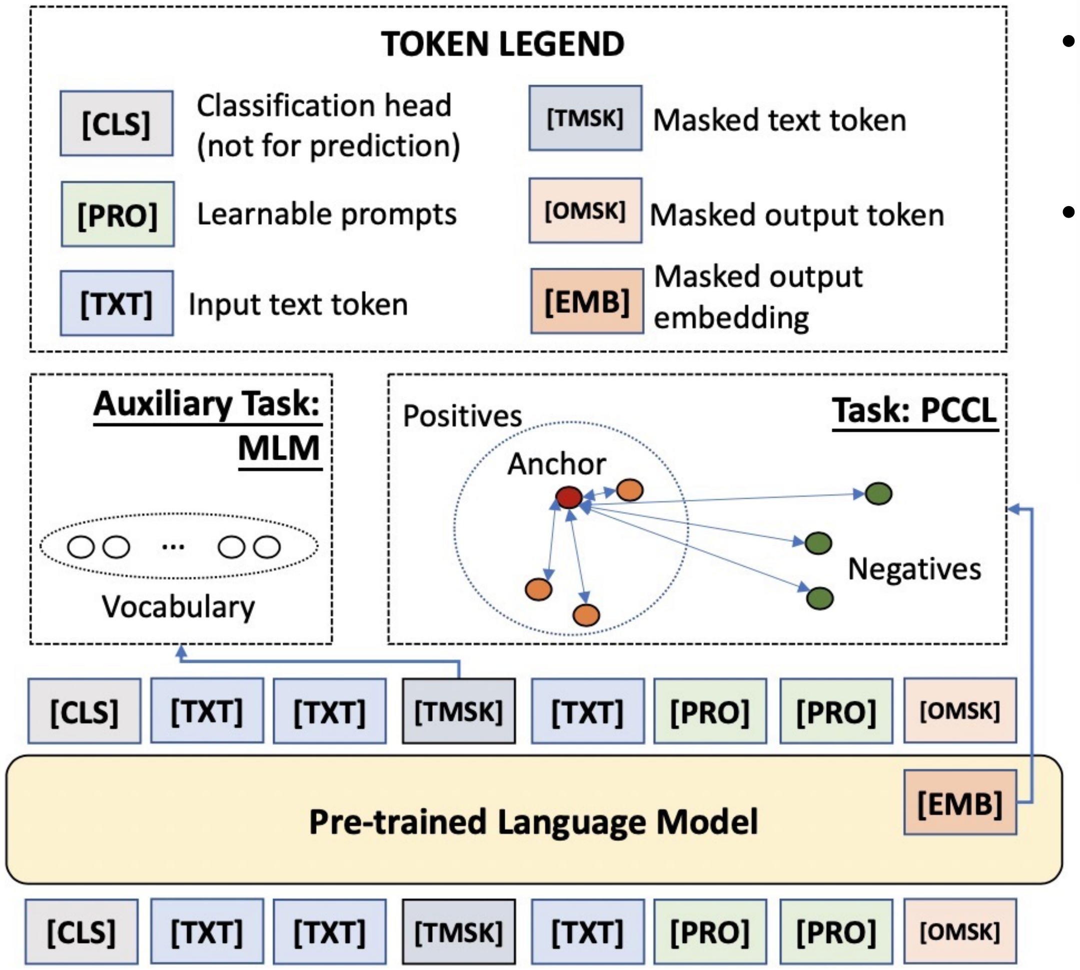


✓ **Current Problems of Prompt-based Fine-tuning**

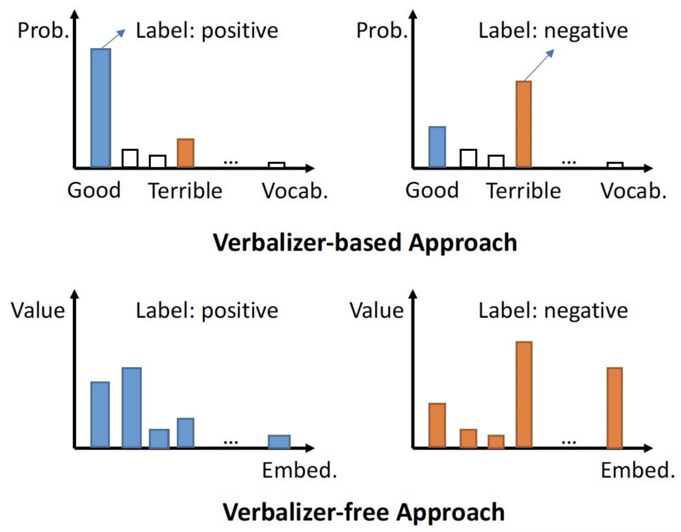
- Manually designed prompts and verbalizers
- Unstable results with different prompts

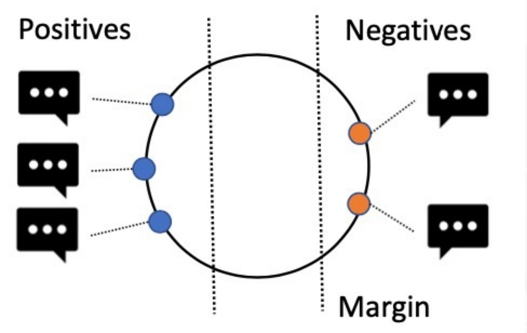
Tianyu Gao, Adam Fisch, Danqi Chen. Making Pre-trained Language Models Better Few-shot Learners. ACL-IJCNLP 2021

Contrastive Prompt Tuning (CP-Turing)



- **Improvement of Prompts**
  - Using continuous prompt embeddings in input
- **Improvement of Verbalizers**
  - Replacing verbalizer mapping with Contrastive Learning





Pairwise Cost-sensitive Contrastive Learning

- **Loss function of CP-Tuning**

$$\mathcal{L}(i) = \mathcal{L}_{PCCL}(i) + \lambda\mathcal{L}_{MLM}(i)$$

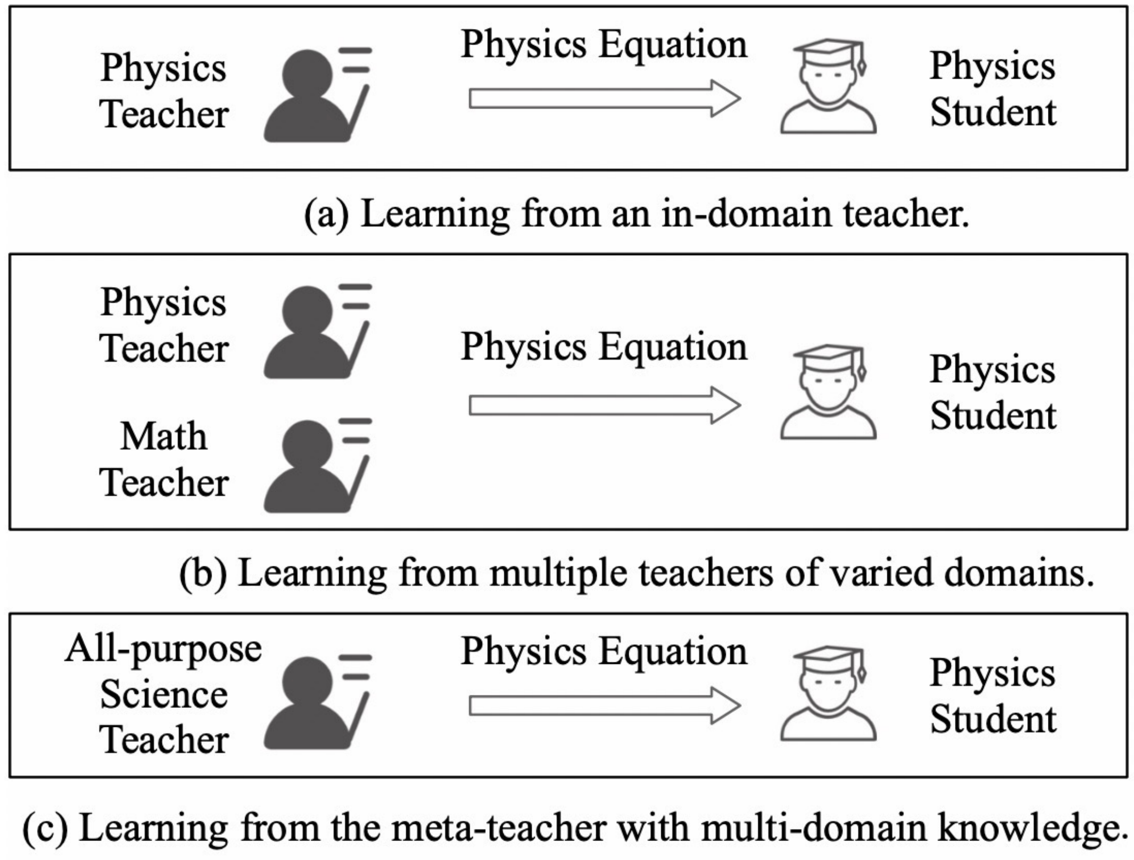
Ziyun Xu*, Chengyu Wang*, Minghui Qiu, Fuli Luo, Runxin Xu, Songfang Huang, Jun Huang. Making Pre-trained Language Models End-to-end Few-shot Learners with Contrastive Prompt Tuning. arXiv

Evaluation Results of CP-Tuning

Backbone	Method	Sentiment Analysis SST-2	Sentiment Analysis MR	Sentiment Analysis CR	Sentence Matching MRPC	Sentence Matching QQP	NLI QNLI	NLI RTE	Subjectivity SUBJ	Avg.
RoBERTa	Standard Fine-tuning	78.62	76.17	72.48	64.40	63.01	62.32	52.28	86.82	69.51
	PET	92.06	87.13	87.13	66.23	70.34	64.38	65.56	91.28	78.01
	LM-BFF (Auto T)	90.60	87.57	90.76	66.72	65.25	68.87	65.99	91.61	78.42
	LM-BFF (Auto L)	90.55	85.51	91.11	67.75	70.92	66.22	66.35	90.48	78.61
	LM-BFF (Auto T+L)	91.42	86.84	90.40	66.81	61.61	61.89	66.79	90.72	77.06
	P-tuning	91.42	87.41	90.90	71.23	66.77	63.42	67.15	89.10	78.43
	WARP	58.80	55.25	55.55	65.74	65.80	52.29	60.07	65.59	59.89
	CP-Tuning	**93.35**	**89.43**	**91.57**	**72.60**	**73.56**	**69.22**	**67.22**	**92.27**	**81.24**
ALBERT	Standard Fine-tuning	63.98	64.90	71.50	56.78	59.32	53.48	52.14	80.54	62.83
	PET	87.11	81.47	88.32	57.21	66.16	55.32	61.85	83.28	72.59
	LM-BFF (Auto T)	82.60	83.23	88.48	**64.04**	60.28	59.42	60.42	84.67	72.75
	LM-BFF (Auto L)	86.83	83.02	89.12	63.43	59.49	56.86	57.33	88.08	73.02
	LM-BFF (Auto T+L)	84.40	82.75	89.52	62.48	56.48	57.69	61.09	88.44	72.85
	P-tuning	85.42	84.32	82.35	58.76	57.46	58.97	55.07	84.32	70.83
	WARP	66.63	65.59	72.34	63.48	58.20	57.45	53.86	62.41	62.49
	CP-Tuning	**89.63**	**84.68**	**90.39**	63.52	**71.05**	**62.02**	**61.92**	**89.02**	**76.52**

Meta Knowledge Distillation (Meta-KD)

- Goal: Improving the effectiveness of knowledge distillation across domains



Analogy Analysis

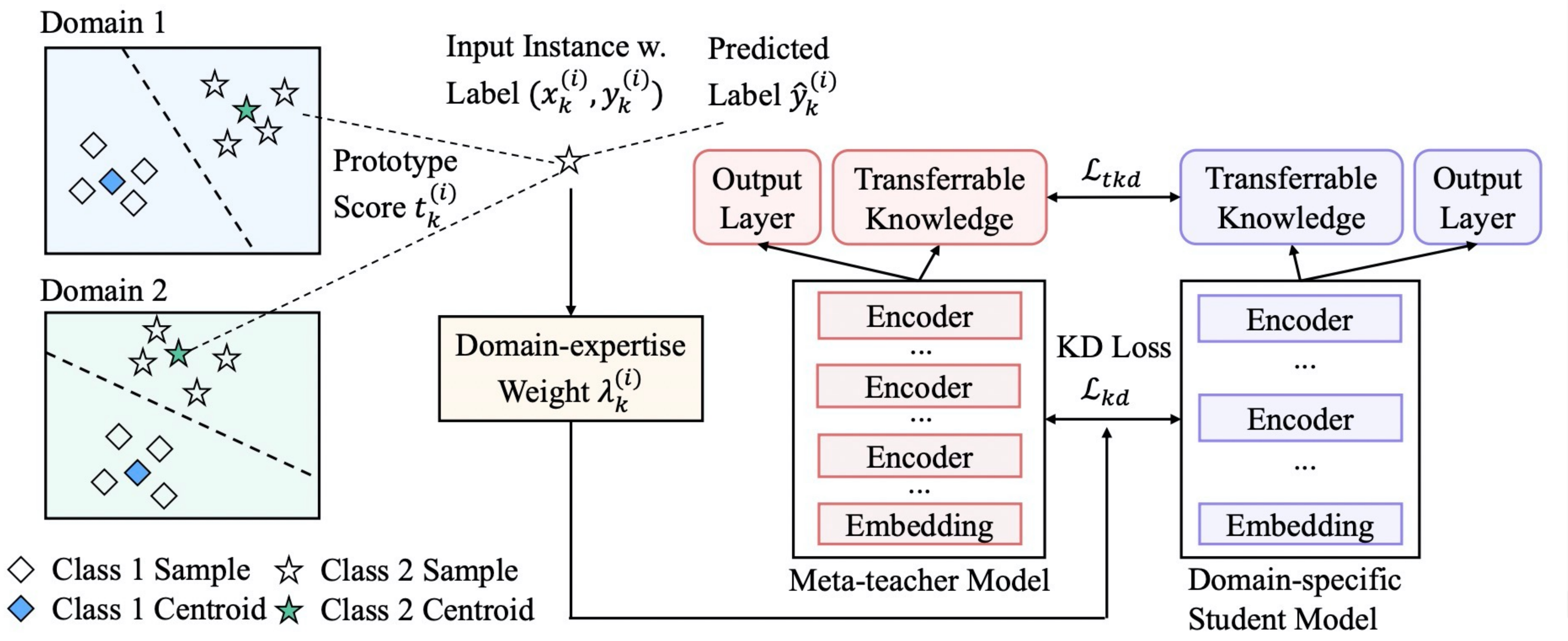
Students who master common knowledge in math and physics can have a better grasp of specific problems in math and physics.

All-purpose Science Teacher -> Meta Leaner

Haojie Pan*, Chengyu Wang*, Minghui Qiu, Yichang Zhang, Yaliang Li, Jun Huang. Meta-KD: A Meta Knowledge Distillation Framework for Language Model Compression across Domains. ACL-IJCNLP 2021

Model Architecture of Meta-KD

- Core idea: **Selectively transferring** cross-domain, transferable knowledge from Meta Teacher to Student



Experimental Results of Meta-KD

- Compared to original BERT, the small model obtained by Meta-KD reduces accuracy by 1.5% only. (#Para. 109M->14.5M)

Methods	Fiction	Government	Slate	Telephone	Travel	Average
BERT-single	82.2	84.2	76.7	82.4	84.2	81.9
BERT-mix	84.8	87.2	80.5	83.8	85.5	84.4
BERT-mtl	83.7	87.1	80.6	83.9	85.8	84.2
Meta-teacher	85.1	86.5	81.0	83.9	85.5	84.4
BERT-single → TinyBERT	78.8	83.2	73.6	78.8	81.9	79.3
BERT-mix → TinyBERT	79.6	83.3	74.8	79.0	81.5	79.6
BERT-mtl → TinyBERT	79.7	83.1	74.2	79.3	82.0	79.7
Multi-teachers → MTN-KD	77.4	81.1	72.2	77.2	78.0	77.2
Meta-teacher → TinyBERT	80.3	83.0	**75.1**	80.2	81.6	80.0
Meta-teacher → Meta-distillation (ours)	**80.5**	**83.7**	75.0	**80.5**	**82.1**	**80.4**

Main Contents

✓Knowledge-enhanced Pre-training

✓Deploying Large Pre-trained Models

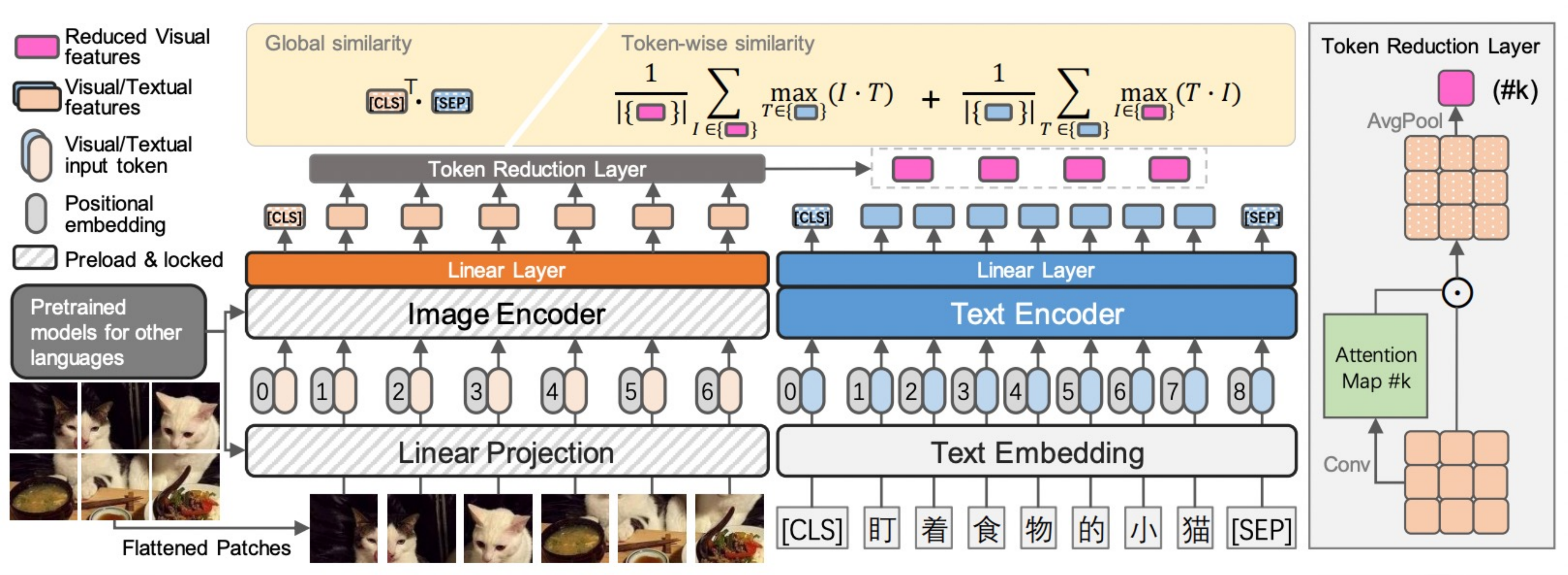
- Prompt-based Few-shot Learning
- Knowledge Distillation for Large Pre-trained Models

✓**Multi-modal Pre-trained Models**

✓Overview of EasyNLP

CLIP-style Models for Text-image Retrieval

✓EasyNLP supports Chinese CLIP-style Models



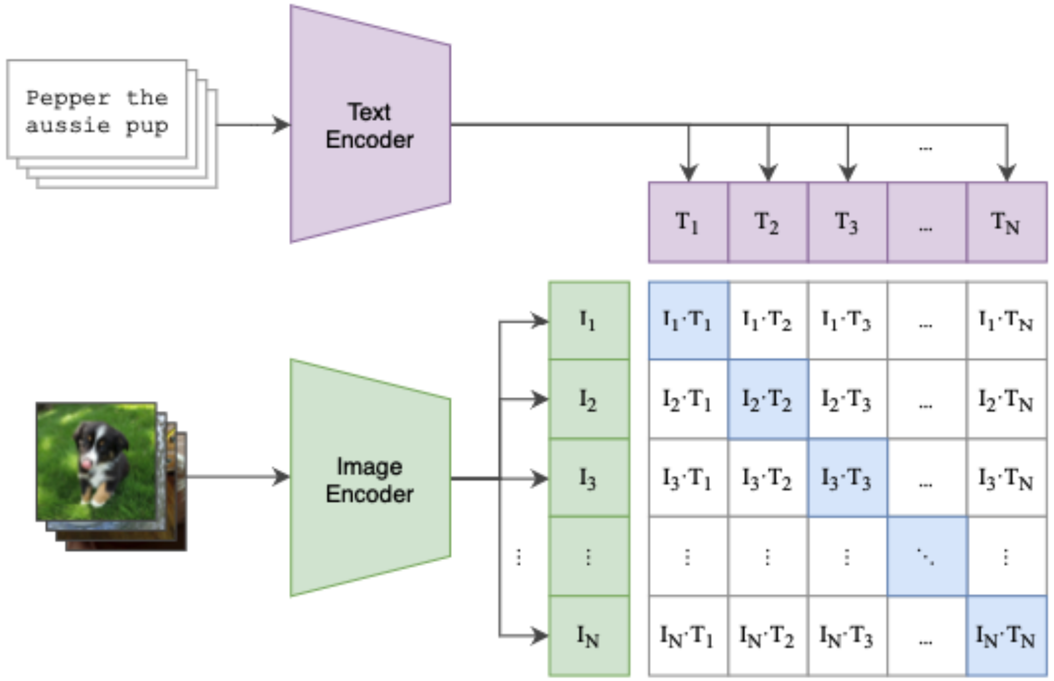
Jiaxi Gu, Xiaojun Meng, Guansong Lu, Lu Hou, Minzhe Niu, Hang Xu, Xiaodan Liang, Wei Zhang, Xin Jiang, Chunjing Xu. Wukong: 100 Million Large-scale Chinese Cross-modal Pre-training Dataset and A Foundation Framework. arXiv

CLIP-style Models for Text-image Retrieval

✓EasyNLP supports SOTA English CLIP-style Models for fashion

Evaluation Results on Fashion-Gen

Model	Rank@1	Rank@5	Rank@10
FashionBERT	26.75	46.48	55.74
KaleidoBERT	33.9	60.5	68.6
CLIP	36.8	58.9	67.6
CommerceMM	39.6	**61.5**	72.7
EI–CLIP	28.4	57.1	69.4
pai–clip–commercial–base–en	39.5	**61.5**	70.0
pai–clip–commercial–large–en	**54.6**	**75.1**	**81.4**

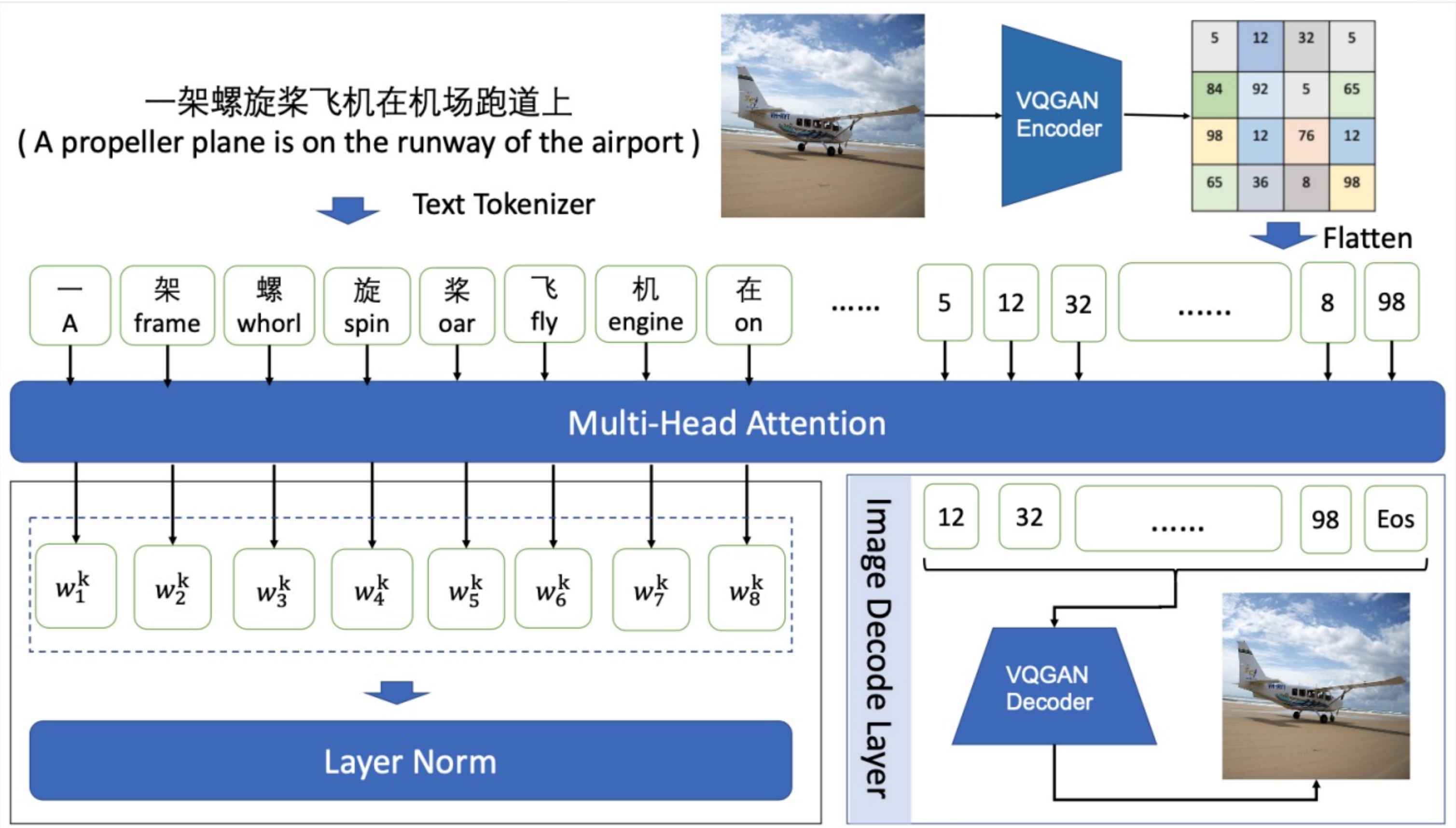


Text	Top-1 result of our CLIP	Top-1 result of OpenCLIP
Canvas slip-on sandals in black. Fringed edges throughout. Open round toe. Leather lining in beige. Round block heel. Tonal leather sole. Tonal stitching. Approx. 3" heel.		
Long sleeve cotton-blend jersey henley in heather 'medium' grey. Crewneck collar. Three-button placket. Rib knit cuffs. Tonal stitching.		
Jersey skirt in black. Elasticized waistband. Shirring at front waist. Drop-tail hem. Fully lined. Tonal stitching.		

DALLE-style Text-to-image Generation

EasyNLP Text-to-image Generation Models

- Specific for the Chinese language
- VQGAN for image generation
- Transformer for converting texts to image tokens
- Moderate model size (<100M parameters)



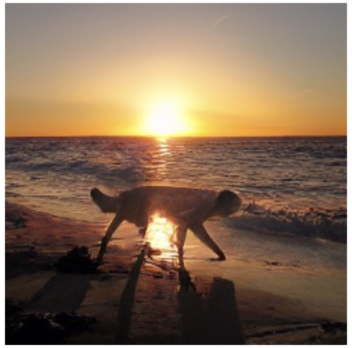
Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, Ilya Sutskever. Zero-Shot Text-to-Image Generation. ICML 2021

一只俏皮的狗正跑过草地
A playful dog is running across the grass




一片水域的景色以日落为背景
A view of water with sunset in the background






Chinese Painting Generation

风阁水帘今在眼，
且来先看早梅红

Red plum blossom

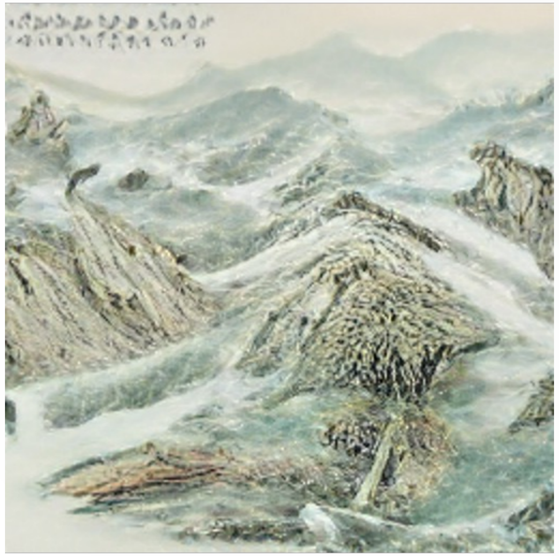
见说春风偏有贺，
露花千朵照庭闹

Thousands of flowers in spring

Chinese Painting Generation

静夜沉沉，
浮光霭霭

Floating mist in quiet night



遥望吴山为谁好，
忽闻楚些令人伤

Seeing mountain view in a sad mood

Main Contents

✓Knowledge-enhanced Pre-training

✓Deploying Large Pre-trained Models

- Prompt-based Few-shot Learning
- Knowledge Distillation for Large Pre-trained Models

✓Multi-modal Pre-trained Models

✓**Overview of EasyNLP**

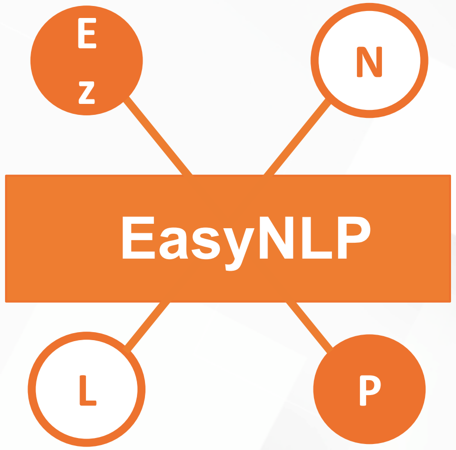
Overview of the EasyNLP Toolkit

✓ History

- In 2021, we started building the EasyNLP toolkit.
- EasyNLP has supported over 10 BUs in Alibaba Group since 2021.
- Starting from May 2022, EasyNLP goes open-sourced in GitHub.

✓ Features of EasyNLP

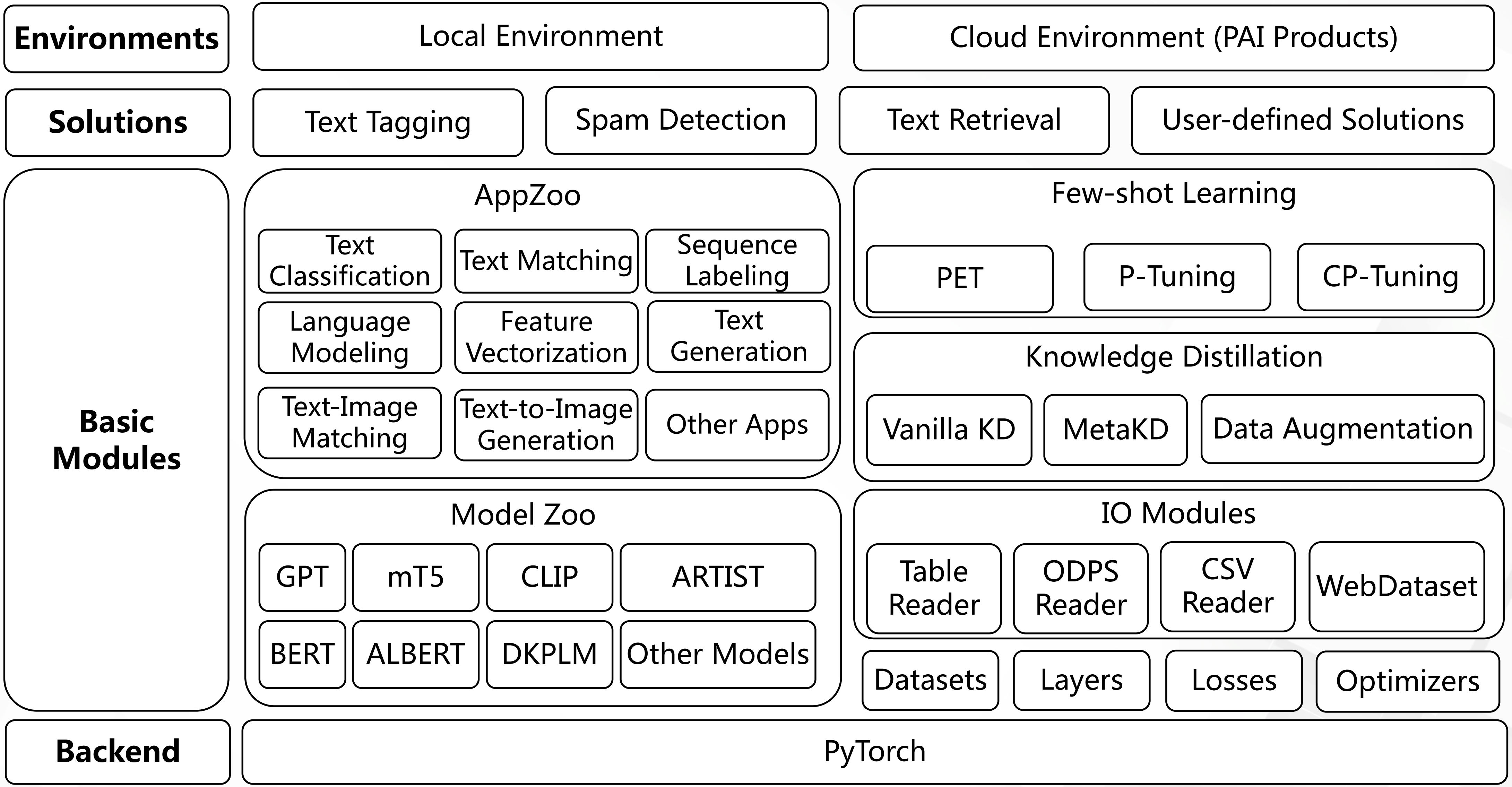
- Easy to use and highly customizable
- Compatible with open-source libraries
- Knowledge-enhanced pre-training
- Deploying large pre-trained models (knowledge distillation and few-shot learning)
- Multi-modal pre-trained models



Alibaba Cloud-PAI

Chengyu Wang, Minghui Qiu, Taolin Zhang, Tingting Liu, Lei Li, Jianing Wang, Ming Wang, Jun Huang, Wei Lin. EasyNLP: A Comprehensive and Easy-to-use Toolkit for Natural Language Processing. EMNLP 2022

EasyNLP Framework



API Examples – AppZoo Mode

✓Black-box commands for model training, evaluation and prediction

Command for text classification

```
easynlp \
    --mode=train \
    --worker_gpu=1 \
    --tables=train.tsv,dev.tsv \
    --input_schema=sent:str:1,label:str:1 \
    --first_sequence=sent \
    --label_name=label \
    --label_enumerate_values=0,1 \
    --checkpoint_dir=./classification_model \
    --epoch_num=1 \
    --sequence_length=128 \
    --app_name=text_classify \
    --user_defined_parameters=‘pretrain_model_name_or_path=bert-small-uncased’
```

API Examples – Python Mode

✓Python APIs for model training, evaluation and prediction

```
from easynlp.dataset import load_dataset, GeneralDataset

# load dataset
dataset = load_dataset('clue', 'tnews')["train"]

# parse data into classification model input
encoded = GeneralDataset(dataset, 'chinese-bert-base')

# load model
model = SequenceClassification('chinese-bert-base')
trainer = Trainer(model, encoded)

# start to train
trainer.train()
```

Load dataset

Data Pre-processing

Load Model

Begin Training

Future Roadmap

✓Knowledge Pre-training

- Releasing more knowledge pre-trained models to improve the models’ understanding abilities of knowledge

✓Multi-modal Pre-training

- Releasing better multi-modal pre-trained models for various tasks

✓Pre-trained Models for Closed-domains

- Supporting various NLP and multi-modal tasks for closed-domains domains (e.g., medicine, finance)

✓Better support for cloud products

- Providing better support on the cloud

Key References

- ✓ Chengyu Wang, Minghui Qiu, Taolin Zhang, Tingting Liu, Lei Li, Jianing Wang, Ming Wang, Jun Huang, Wei Lin. EasyNLP: A Comprehensive and Easy-to-use Toolkit for Natural Language Processing. EMNLP 2022
- ✓ Taolin Zhang*, Chengyu Wang*, Nan Hu, Minghui Qiu, Chengguang Tang, Xiaofeng He, Jun Huang. DKPLM: Decomposable Knowledge-enhanced Pre-trained Language Model for Natural Language Understanding. AAAI 2022
- ✓ Tianyu Gao, Adam Fisch, Danqi Chen. Making Pre-trained Language Models Better Few-shot Learners. ACL-IJCNLP 2021
- ✓ Ziyun Xu*, Chengyu Wang*, Minghui Qiu, Fuli Luo, Runxin Xu, Songfang Huang, Jun Huang. Making Pre-trained Language Models End-to-end Few-shot Learners with Contrastive Prompt Tuning. arXiv
- ✓ Haojie Pan*, Chengyu Wang*, Minghui Qiu, Yichang Zhang, Yaliang Li, Jun Huang. Meta-KD: A Meta Knowledge Distillation Framework for Language Model Compression across Domains. ACL-IJCNLP 2021
- ✓ Jiaxi Gu, Xiaojun Meng, Guansong Lu, Lu Hou, Minzhe Niu, Hang Xu, Xiaodan Liang, Wei Zhang, Xin Jiang, Chunjing Xu. Wukong: 100 Million Large-scale Chinese Cross-modal Pre-training Dataset and A Foundation Framework. arXiv
- ✓ Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, Ilya Sutskever. Zero-Shot Text-to-Image Generation. ICML 2021

Following Our GitHub Project

Ez N EasyNLP L P

https://github.com/alibaba/EasyNLP

EasyNLP开源用户群 (含 ...
428人

扫一扫群二维码，立刻加入该群。

THANKS

--------- Q&A Section --------