澜舟科技
langboat

孟子轻量化预训练模型

SMP10周年纪念活动之大模型技术研讨会

主讲人：
周明、张倬胜

周明：创新工场首席科学家、澜舟科技创始人
中国计算机学会副理事长、国际计算语言学会原主席

张倬胜：上海交大博士生（导师：赵海教授）
入选2021全球AI华人新星百强、澜舟科技实习生

Part . 01

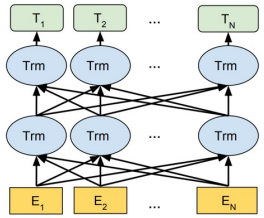
01 孟子模型

大规模预训练模型

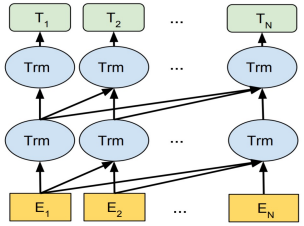
用一套机制（预训练+微调）解决所有语言、主要场景的NLP任务，解决了原有的碎片化问题，大大提升开发效率。标志着NLP进入工业化实施阶段。

技术路线

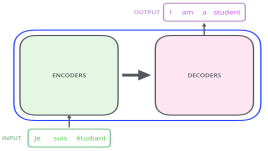
主要模型



Encoder(BERT-Style)



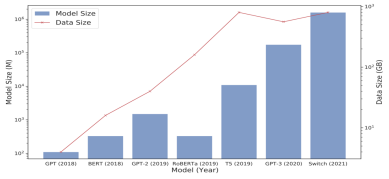
Decoder(GPT-Style)



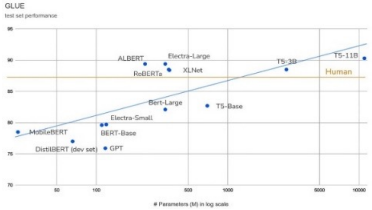
Encoder-Decoder
(T5-Style)

发展趋势

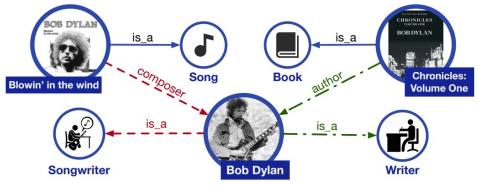
趋势 1：更大的模型和更多的数据



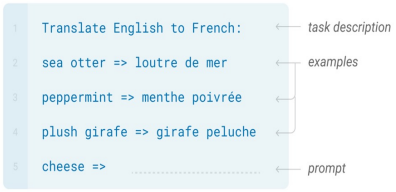
趋势 2：更高效的预训练方法



趋势 3：知识增强的预训练模型



趋势 4：小样本学习及统一微调

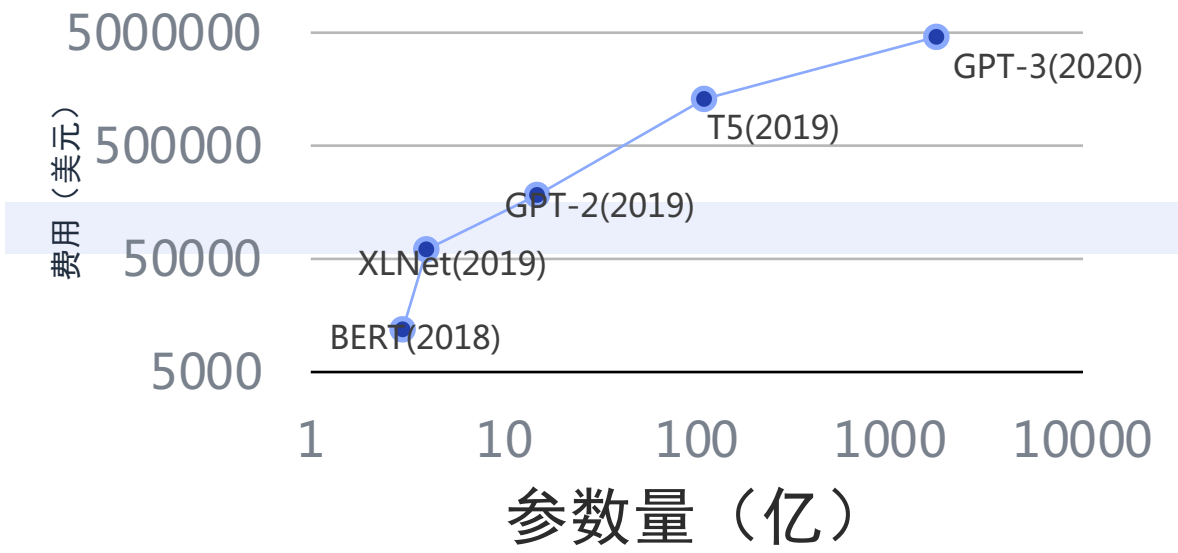


为什么要训练轻量化模型？

2018年到2020年3年间预训练模型的**参数量增长了3个数量级**，而硬件能力提升速度远低于模型参数量增长速度，训练费用仍然**上升了2个数量级。**

公开渠道统计的模型训练成本

费用计算年份	模型	参数量（十亿）	费用（美元）
2018	BERT	0.34	12,000
2019	XLNet	0.34	61,000
2019	GPT-2	1.5	184,320
2019	T5	11	1,300,000
2020	GPT-3	175	4,600,000



训练价格的影响因素包括：

1. 模型参数量
2. GPU/TPU算力
3. 数据量

模型参数量与训练所需数据量的爆发性增长导致训练费用显著增长。

在实际任务应用时，巨型模型面临与下游任务不能灵活适配、实施代价大、不节能环保等严重问题。

有鉴于此，我们需要降低训练成本。比如，提升训练能力，加快训练速度，研究轻量型模型。

孟子轻量化预训练模型

澜舟科技 langboat

CLUE总排行旁记最佳得分 (10.22后中文原版数据集OCNLI.bert-base初始化分数,可重新跑并提交)

模型参数	排名	模型	研究机构	测评时间	总分	认证	AFQMC 语义相似度	TNEWS 短文本分类	IFLYTEK 长文本分类	CMNLI 自然语言推理	CMNLI-50K 自然语言推理	CLUEWSC20 20代词消歧	CSJ 关键词识别	CMRC2018 阅读理解	CHID 成语填空	C3 阅读理解2
	1	HUMAN	CLUE	19-12-01	85.610	已认证	81.000	71.000	80.300	76.000	90.3	98.000	84.000	92.400	87.100	96.000
十亿	2	Mengzi	澜舟科技-创新工场	21-07-12	84.261	待认证	79.82..	75.06	65.07..	86.1295	81.867	96.5517	89.86..	82.250	96.002	89.979
百亿	3	Motion	QQ浏览器搜索	21-06-25	84.055	待认证	78.29..	73.18	65.46..	85.4374	84.967	94.8275	90.16	85.300	94.425	88.489
百亿	4	BERTSG	Sogou search	21-06-25	83.824	待认证	79.84..	74.15	64.53..	85.2990	85.933	95.1724	89	83.800	93.059	87.436
千亿	5	Pangu	华为云-循环智能	21-04-23	83.046	待认证	78.114	72.070	65.192	85.190	83.00	95.517	87.733	84.450	93.253	85.637

小

100M 至 1B 参数量多级别模型针对不同需求。低硬件需求，低研发成本。

精

模型结构上引入更多知识，同样模型体积下更好的表现。

快

8 张 3090 约 3 天完成一个领域迁移（base 级）8 张 3090 半天完成一个任务适应 。

专

每个领域每个任务定制预训练模型，超过通用的大模型。

合作伙伴: 知乎 有问题 上知乎 · 上海交通大學 SHANGHAI JIAO TONG UNIVERSITY · 北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY · 北京大學 PEKING UNIVERSITY · 北京交通大學

扫码加入孟子开源社区微信群

直接访问开源项目地址
https://github.com/Langboat/Mengzi

访问澜舟科技官方网站
https://langboat.com

澜舟科技
langboat

澜舟科技已将孟子中文预训练系列模型开源！

全面覆盖文本理解、文本生成、金融、多模态等不同技术应用领域。希望与开源社区一起让新一代文本智能技术落地各行各业，推动下一波认知智能生产力进化的浪潮。

开源孟子轻量化预训练模型

主题：同等规模下更强的性能，良好的兼容性和易用性

应用：涵盖文本理解，文本生成，垂直金融和多模态等场景

模型	参数量	特点	适用任务	语料
Mengzi-BERT-base	110M	兼容 BERT 架构，利用语言学知识增强模型能力	文本分类、实体识别、关系抽取等	300G 互联网语料
Mengzi-BERT-base-fin	110M	基于 Mengzi-BERT-base 在金融语料上继续训练	金融新闻分类、信息抽取、情感分析	+20G 金融新闻、公告、研报
Mengzi-T5-base	220M	可以提升文本生成的可控性，优于 GPT 结构	文案生成、新闻生成等	300G 互联网语料
Mengzi-Oscar-base	110M	基于 Mengzi-BERT-base，在百万级图文对上进行训练	图片描述、图文互检等	百万级图文对

相关文档以及模型下载： https://github.com/Langboat/Mengzi

Zhuosheng Zhang, Hanqing Zhang, Keming Chen, Yuhang Guo, Jingyun Hua, Yulong Wang, Ming Zhou.
Mengzi: Towards Lightweight yet Ingenious Pre-trained Models for Chinese. https://arxiv.org/abs/2110.06696.

开源模型下游任务表现

CLUE 任务

模型	AFQMC	TNEWS	IFLYTEK	CMNLI	WSC	CSL	CMRC	C3	CHID
RoBERTa-wwm-ext (baseline)	74.30	57.51	**60.80**	80.70	67.20	80.67	77.59	67.06	83.78
Mengzi-BERT-base	**74.58**	**57.97**	60.68	**82.12**	**87.50**	**85.40**	**78.54**	**71.70**	**84.16**

金融领域任务

模型	检索 (Recall@10/20)	实体识别	关系抽取	实体链指
RoBERTa-wwm-ext (baseline)	90.20/92.90	88.11	77.44	93.40
Mengzi-BERT-base	90.40/92.40	88.51	77.51	93.80
Mengzi-BERT-base-fin	**91.00/93.50**	**88.53**	**77.57**	**94.10**

图片描述

Microsoft Office 自动替换文字:

人骑着马

Mengzi-Oscar:

绿油油的草地上有两个面带微笑的人在骑马

Microsoft Office 自动替换文字:

粉色的伞走在路上的小孩

Mengzi-Oscar:

两个打着伞的人和一个背着孩子的男人走在被水淹没的道路上

孟子知识服务引擎

轻量化预训练模型

基于自研的孟子轻量化预训练语言模型，处理多语言和多模态数据，支持理解和生成。通过订制满足不同领域、不同应用场景的需求。荣登中文NLP评测CLUE的榜首。

文本生成

采用预训练语言模型、通用和领域大数据，开发交互式可控文本生成技术，指定关键词、知识单元、应用场景生成文本。用于营销文案生成、新闻摘要、小说和剧本创作。

机器翻译

采用预训练语言模型和多语言联合训练、术语识别等技术，实现以中文为中心的世界主要语言之间的互译。为金融、工程、制造等垂直领域打造专用翻译引擎。

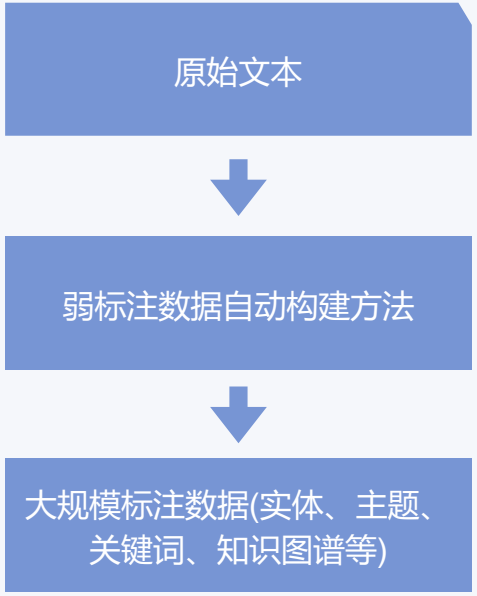
搜索引擎

基于预训练模型和知识图谱、开发新一代知识服务引擎。汇总海量实时信息，提供检索、问答、文摘、洞见，提升行业全流程效率。用于金融、营销、法律、政务等领域。

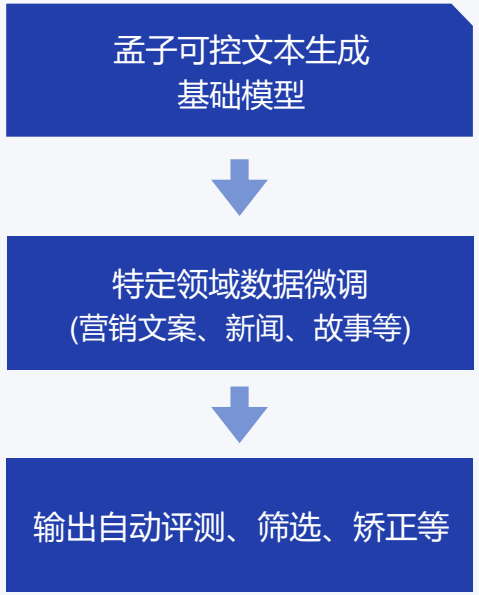
孟子模型应用示例：澜舟可控文本生成

- 基于孟子预训练模型，采用预训练+微调的学习方式。
- 构建条件扩写预训练任务，为给定输入关键信息，生成对应的目标文本（段落、篇章等）。
- 针对特定应用场景，构建学习数据和进行快速领域适配。
- 输入内容的自动推荐，提高生成的多样性和准确性。

大规模训练数据自动构建



文本生成引擎构建流程



可控文本生成预训练任务

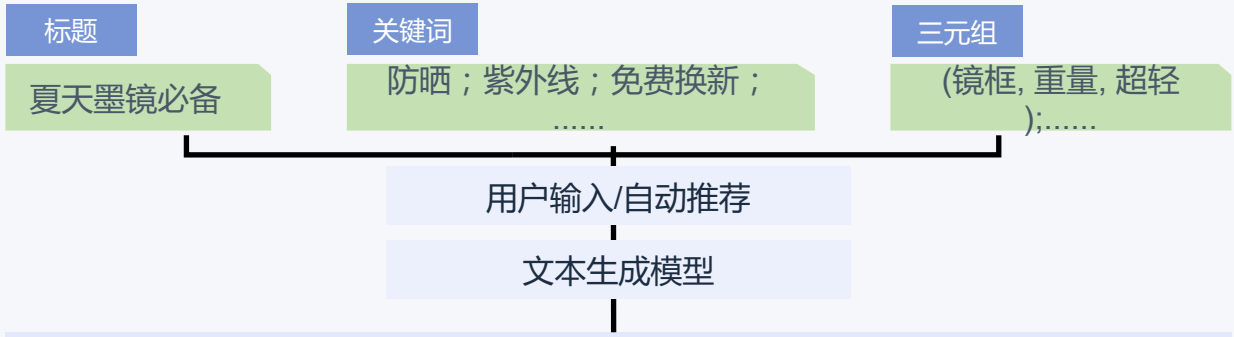
输入

春天已经来了，夏天还会远吗？<X>|||夏天、防晒、紫外线等信息

输出

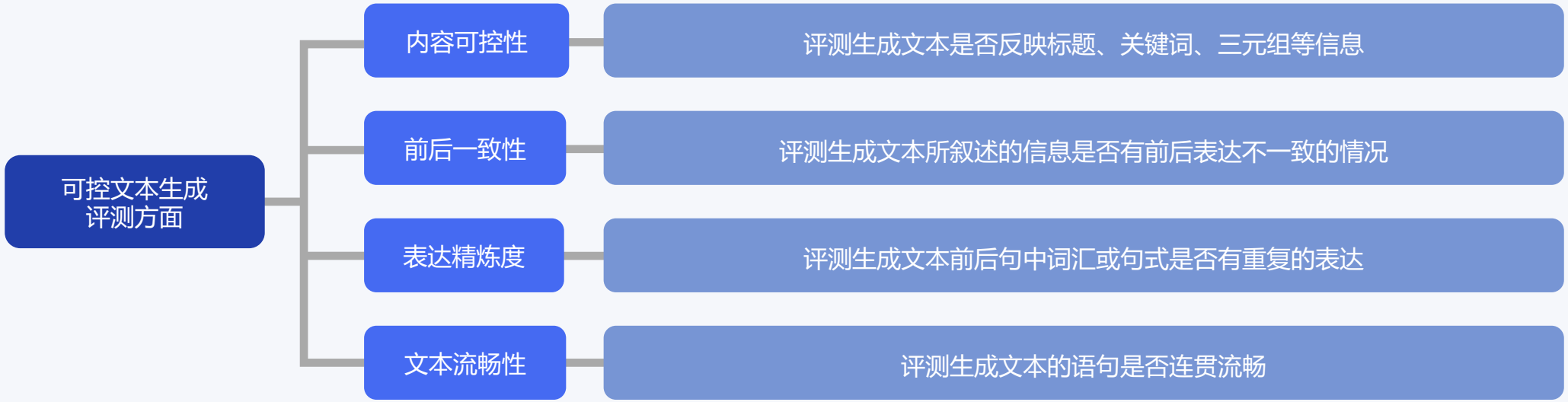
<X>每到夏天来临之际，我就会给闺女准备各种防晒必需品，毕竟紫外线太强了.....

可控文本生成应用示例



夏天马上就要到了，紫外线也越来越强，防晒工作一定要做好，今天给大家推荐一款我最近超级喜欢的墨镜，就是这款genanx太阳镜。镜片材质pc，质感柔软，镜框重量超轻，戴上去很轻盈，不会有压迫感，而且还可以折叠收纳，出门携带很方便。它的镜腿是加宽加厚的设计，能够很好的遮挡住我们脸上的一些小瑕疵，这样就不用担心会被晒黑了。还有一个很重要的功能，它是防uva和uvb双重防护的.....

文本生成评估体系



	内容可控性检测	前后一致性检测	表达精炼度检测
输入	标题：洗发水 关键词：姜汁，生活，饮食、护理 三元组：<脱发，困扰，男性>	奥巴马昨天在北京访问。奥巴马然后在华盛顿对中国学生发表演讲	夏天适合穿什么裙子呢?夏天到了，又到了穿裙子的季节。清凉大方的红裙子，让你在炎热的夏天也能穿出美丽的气质。裙子颜色是红色，材质是麻布，清凉透气，穿起来很美丽。
输出	脱发是很多女性的烦恼, 很多女性都深受脱发的困扰,生活压力大,饮食不规律,没有规律的护理,很容易出现脱发的情况。今天给大家推荐一款姜汁洗发水,让生活更健康。	小明非常喜欢冲浪。所以他经常到海边玩，其实小明不喜欢冲浪	

可控文本生成结果评测

孟子模型中不同输入的性能评测

孟子基础模型	B-1↑	B-2↑	B-4↑	R-1↑	R-2↑	Dist-1↑	Dist-2↑	Conciseness↑	Controllability↑	PPL↓
标准答案	100	100	100	100	100	79.7	95.2	45.4	100	43.0
输入标题	20.5	7.0	4.3	23.2	7.9	77.2	93.7	32.1	75.0	30.1
输入标题+关键词	31.2	16.1	11.9	36.3	18.7	75.5	92.7	32.5	80.3	28.3
输入标题+三元组	28.2	15.3	13.2	32.2	17.4	77.1	93.3	34.7	81.6	27.9
输入标题+关键词+三元组	37.7	23.4	19.5	43.6	27.0	77.4	93.7	35.6	82.1	27.7

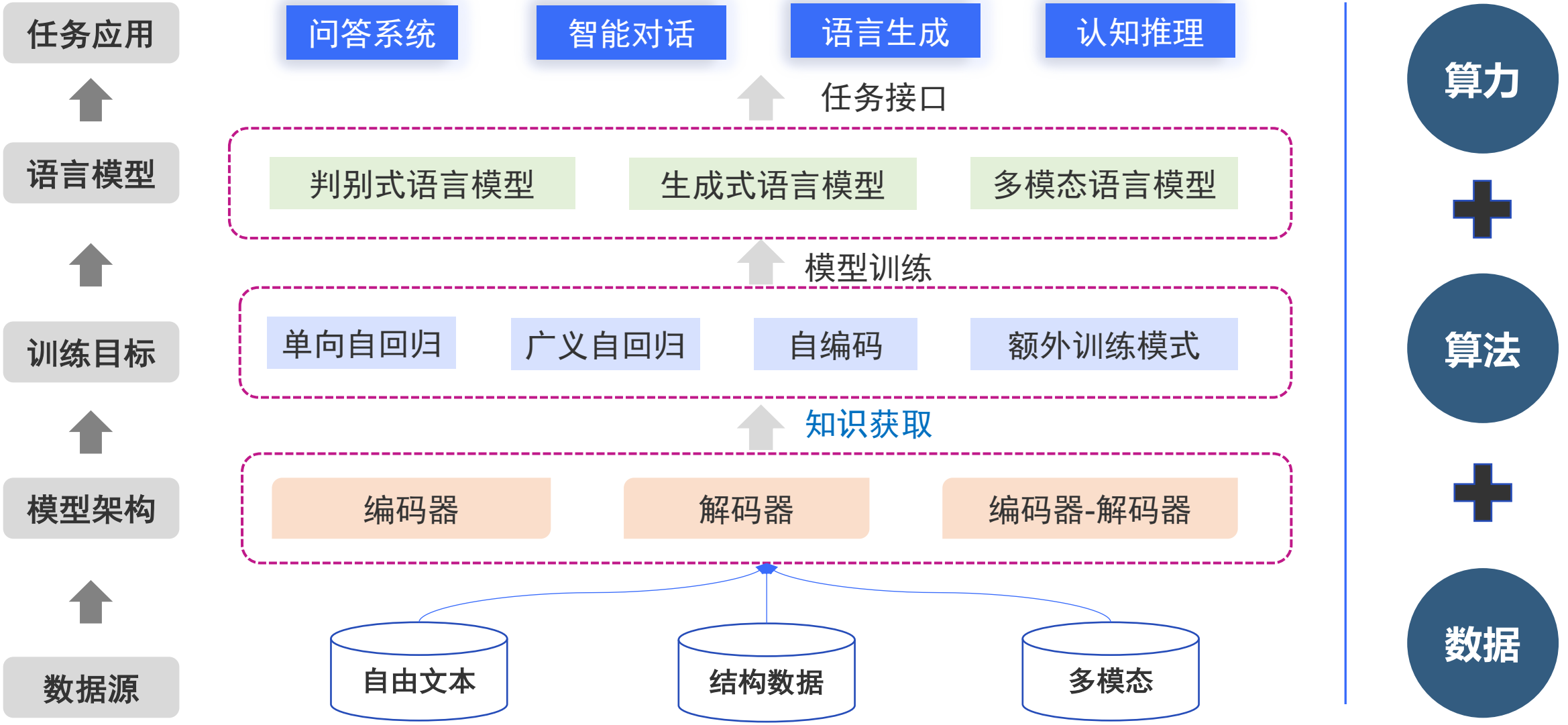
不同预训练模型的性能评测

模型	B-1↑	B-2↑	B-4↑	R-1↑	R-2↑	Dist-1↑	Dist-2↑	Conciseness↑	Controllability↑	PPL↓
孟子基础模型	37.7	23.4	19.5	43.6	27.0	77.4	93.7	35.6	82.1	27.7
孟子可控文本生成模型	39.3	24.6	20.5	44.7	27.7	78.1	94.4	36.9	86.0	26.8
GPT	37.1	23.0	19.2	43.3	27.1	78.0	94.2	36.8	56.0	26.1

Part . 02

02 轻量化训练技术

轻量化模型研究框架

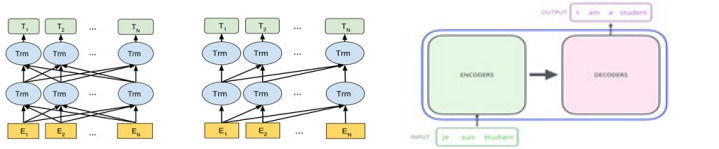

任务应用
语言模型
训练目标
模型架构
数据源
问答系统
智能对话
语言生成
认知推理
任务接口
判别式语言模型
生成式语言模型
多模态语言模型
模型训练
单向自回归
广义自回归
自编码
额外训练模式
知识获取
编码器
解码器
编码器-解码器
自由文本
结构数据
多模态
算力
算法
数据

轻量化模型研究

预训练概况

预训练模型类型：

- 自回归模型：如GPT
- 自编码模型：如BERT (MLM)
- Encoder-Decoder模型：T5



(a) Encoder (b) Decoder (c) Encoder-Decoder

迈向轻量化预训练

为什么要轻量化？

- 训练费用的显著增长
- 硬件能力跟不上模型规模增长
- 巨型模型难以与下游任务适配

何为轻量化预训练：现实应用为导向

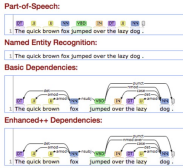
- 同等规模下性能更强、效率更高、更鲁棒

知识增强

让模型学习更丰富的知识

知识图谱增强

- 融合实体信息的表示
- 基于图谱的推理强化

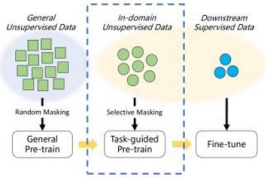


基于语言学知识增强

成分句法 → 基于成分的预训练方法

语义角色 → 语言表示融合增强

依存关系 → 自注意力权重约束和剪枝



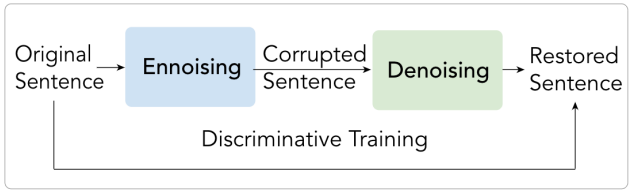
特定数据增强

- 领域相关知识
- 任务相关知识
- 多语言知识

训练优化

让模型更高效地学习知识

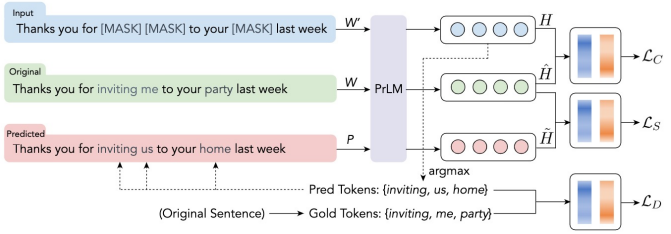
- 模型尺寸优化，减少结构冗余
- 新的训练目标，加强语义学习能力



☹ A cute [MASK] is [MASK] on the [MASK] ...

🙂 [MASK] cute dog [MASK] playing on the [MASK] ...

- 训练校正策略，实现更准确的训练
- 对抗样本预训练，提升模型鲁棒性
- 训练加速策略，分布式大批次训练



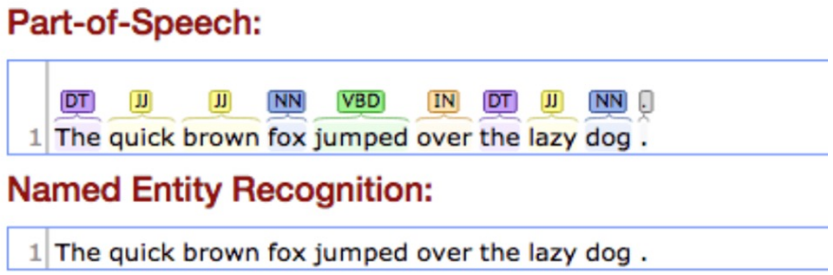
Model： https://github.com/Langboat/Mengzi

Report: Zhang Zhuosheng, et al. Mengzi: Towards Lightweight yet Ingenious Pre-trained Models for Chinese. https://arxiv.org/abs/2110.06696.

轻量化技术-语言学知识增强

语言学信息增强

- 使用SpaCy进行词性标注(POS)和命名实体识别(NER)
- 将识别的目标标签作为预测目标用于训练
- 将POS与NER的预测损失与原始语言建模Loss相加得到最终损失



Model	AVG	AFQMC	TNEWS	IFLYTEK	WSC	CSL	OCNLI	NER	CMRC
RoBERTa-base	73.57	74.44	57.07	59.33	83.22	80.13	76.92	79.69	87.45/68.13
Mengzi Base 0809	73.87	74.61	57.92	60.94	83.55	80.70	76.14	79.56	87.39/67.75
Mengzi Base MultiTask Exp 0830	74.02	73.73	58.02	61.02	84.87	80.77	76.47	79.81	87.22/67.66
RoBERTa-large	75.76	75.42	59.13	61.45	87.83	82.60	78.37	79.94	89.78/72.82
Mengzi Large 0830	76.30	75.28	58.57	61.33	91.12	83.80	79.19	79.97	90.03/72.32

Junru Zhou, Zhuosheng Zhang, Hai Zhao, Shuailiang Zhang. LIMIT-BERT : Linguistic Informed Multi-Task BERT. EMNLP 2020 Findings.

轻量化技术-序列关系增强

序列关系增强

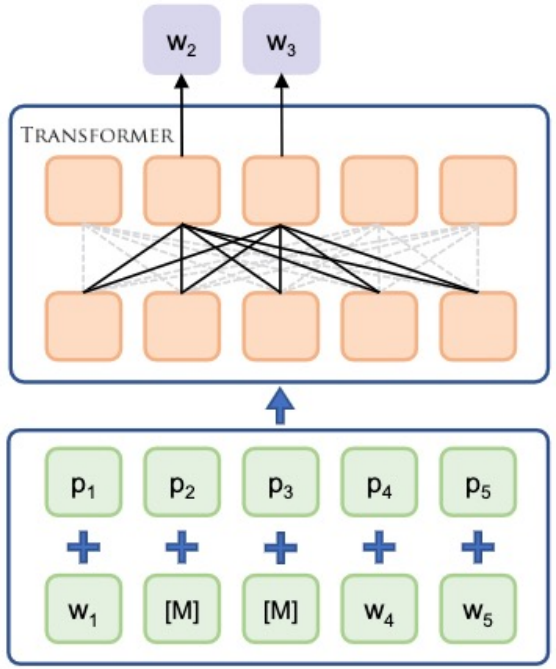
- 结合Sentence Order Prediction (SOP) 训练任务 （有效）
- 结合Next Sentence Prediction (NSP) 训练任务 （不明显）

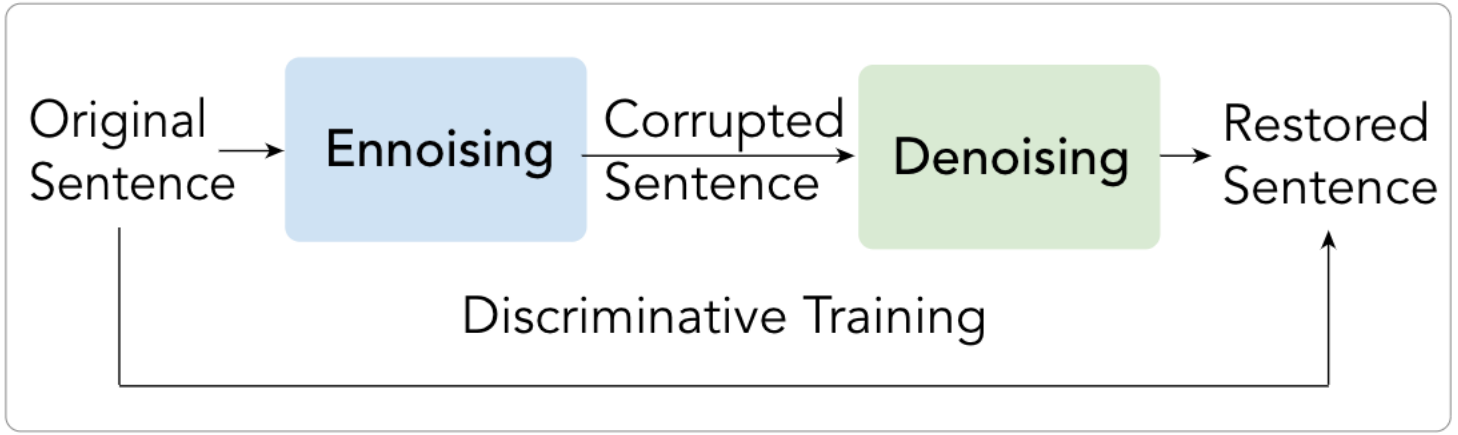
Model	AVG	AFQMC	TNEWS	IFLYTEK	WSC	CSL	OCNLI	NER	CMRC
RoBERTa-base (CLUE Report)	-	74.04	56.94	60.31	~~67.80~~	81.00	-	-	87.28/67.89
RoBERTa-base (Our re-run with CLUE baseline code)	73.57	74.44	57.07	59.33	83.22	80.13	76.92	79.69	87.45/68.13
Mengzi Bert Base 0809	73.87	74.61	57.92	60.94	83.55	80.70	76.14	79.56	87.39/67.75
Mengzi Bert Base SOP 200K	73.97	74.10	57.46	60.98	82.89	81.07	76.37	80.32	88.32/69.01
Mengzi Bert Base SOP 200K hp opt	**74.68(+1.11)**	**75.02(+0.98)**	**57.58(+0.51)**	**60.98(+1.65)**	**87.50(+4.28)**	**81.07(+0.94)**	**76.37(-0.55)**	**80.32(+0.63)**	**88.32(+0.87)/69.01(+0.88)**

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, Radu Soricut. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. ICLR 2020.

基于掩码的预训练：

- **两阶段过程：**构造训练样本（Ennoising），还原被破坏的句子（Denoising）
- **样本结构破坏：**Ennoising会对句子结构造成破坏，样本预测难易度不同。如何根据样本难度自适应学习强度？
- **假负预测问题：**模型还原成与原句不同但合法的句子，会被判定为错。如何在有限的样本下提升训练准确性？



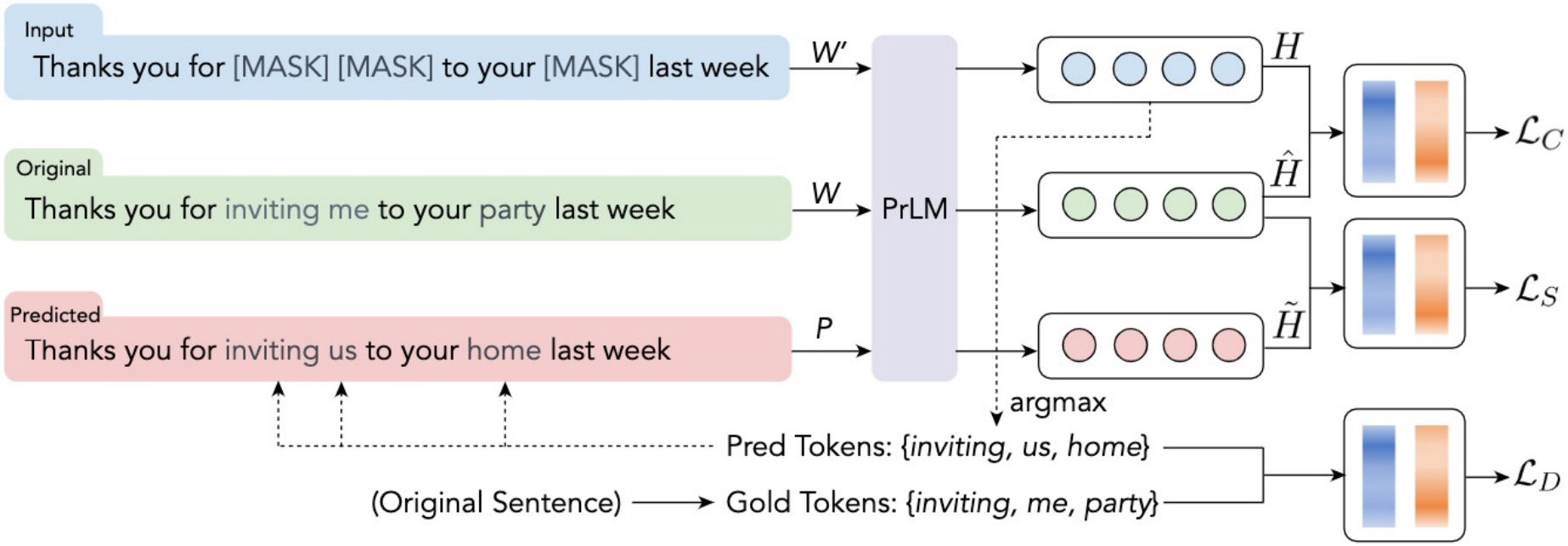


A cute [MASK] is [MASK] on the [MASK] ...

[MASK] cute dog [MASK] playing on the [MASK] ...

训练校正策略 (Noising Potency Estimation)：

- **结构破坏度评估**：计算MASK后的序列与原始序列的分布相似性
- **语义近似度评估**：计算预测出来的序列与原始序列的分布相似性
- 将以上相似性作为训练惩罚项



$$\mathcal{L}_H = D_{KL}(H||\hat{H}) = \sum_{i=1}^{c} H_i \log(\frac{H_i}{\hat{H}_i}) \quad \mathcal{L}_S = D_{KL}(\tilde{H}||\hat{H}) = \sum_{i=1}^{c} \tilde{H}_i \log(\frac{\tilde{H}_i}{\hat{H}_i})$$

训练校正策略 (Noising Potency Estimation)：

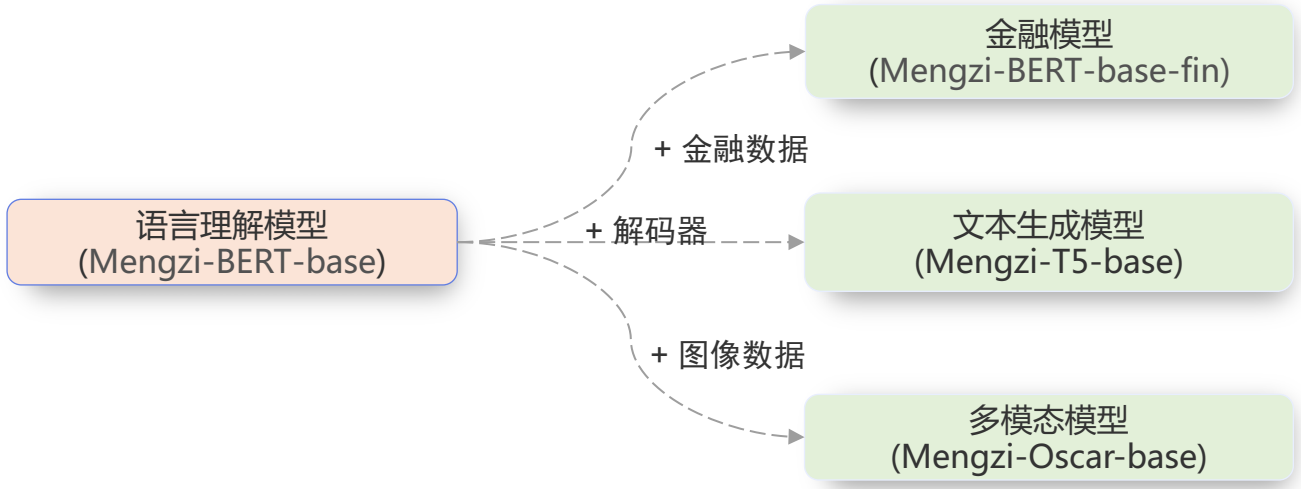
- 有效提升模型在下游任务上的性能
- 有助于模型应对同义词替换的对抗攻击

Model	CoLA *Mcc*	SST-2 *Acc*	MRPC *Acc*	STS-B *Spear*	QQP *Acc*	MNLI *Acc*	QNLI *Acc*	RTE *Acc*	Average -
$\text{BERT}_{\text{base}}$	59.32	92.32	87.25	87.36	90.78	84.75	91.42	65.34	82.32
$\text{BERT-NPE}_{\text{base}}$	61.39	93.46	87.50	89.05	90.90	85.28	91.84	68.95	83.52
$\text{BERT}_{\text{large}}$	62.45	93.58	88.24	90.48	91.45	87.20	92.37	74.01	84.97
$\text{BERT-NPE}_{\text{large}}$	64.07	94.27	88.73	90.57	91.55	87.35	92.71	75.09	85.54
$\text{ELECTRA}_{\text{base}}$	65.53	94.95	88.97	89.96	91.24	88.45	92.53	77.62	86.16
$\text{ELECTRA-NPE}_{\text{base}}$	68.95	95.30	90.44	90.52	91.40	88.66	93.04	79.06	87.17
$\text{ELECTRA}_{\text{large}}$	70.41	96.79	89.22	91.92	92.07	90.26	94.40	85.92	88.87
$\text{ELECTRA-NPE}_{\text{large}}$	72.09	97.48	91.18	92.03	92.27	90.55	94.64	87.36	89.70

Model	*Original Reference*		*SwapSynWordEmbedding*		*SwapSynWordNet*	
	EM Score	F1 Score	EM Score	F1 Score	EM Score	F1 Score
$\text{BERT}_{\text{base}}$	85.33	88.78	84.67 (↓0.67)	87.67 (↓1.11)	81.67 (↓3.67)	85.15 (↓3.63)
$\text{BERT-EPE}_{\text{base}}$	84.33	87.70	84.67 (↑**0.33**)	87.82 (↑**0.12**)	82.33 (↓**2.00**)	85.42 (↓**2.28**)
$\text{ELECTRA}_{\text{base}}$	89.00	90.91	86.67 (↓2.33)	88.89 (↓2.02)	87.00 (↓2.00)	89.39 (↓1.53)
$\text{ELECTRA-EPE}_{\text{base}}$	89.67	91.44	89.00 (↓**0.67**)	90.30 (↓**1.14**)	89.00 (↓**0.67**)	91.03 (↓**0.41**)

孟子语言理解模型(Mengzi-BERT)

- 可微调用于语言编码和理解任务
- 也用于初始化三个模型的语言编码



模型	参数量	特点	适用任务	语料
Mengzi-BERT-base	110M	兼容 BERT 架构，利用语言学知识增强模型能力	文本分类、实体识别、关系抽取等	300G 互联网语料
Mengzi-BERT-base-fin	110M	基于 Mengzi-BERT-base 在金融语料上继续训练	金融新闻分类、信息抽取、情感分析	+20G 金融新闻、公告、研报
Mengzi-T5-base	220M	可以提升文本生成的可控性，优于 GPT 结构	文案生成、新闻生成等	300G 互联网语料
Mengzi-Oscar-base	110M	基于 Mengzi-BERT-base，在百万级图文对上进行训练	图片描述、图文互检等	百万级图文对

孟子多模态模型（基于Oscar+）

中文多模态研究的挑战

- 缺乏中文大规模图文对齐的预训练数据
- 最佳的模型架构依然还没有定论

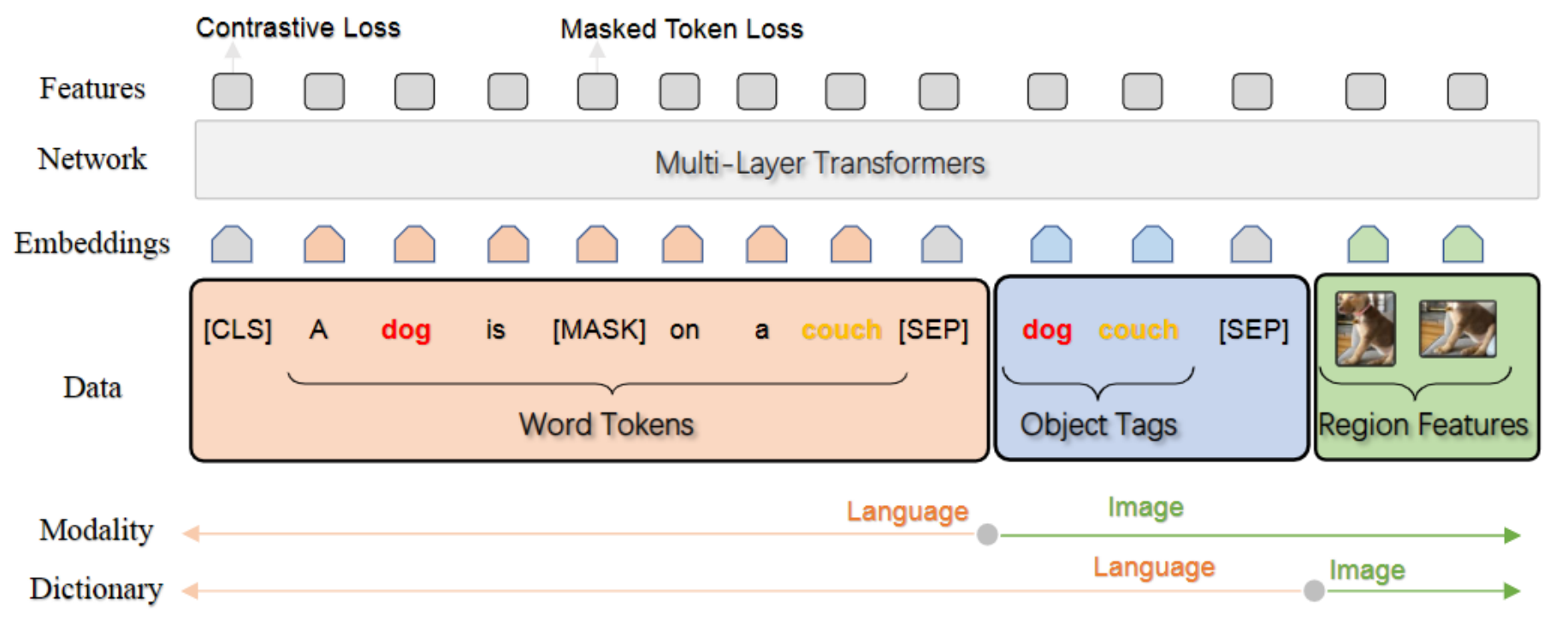
解决方案

- 通过机器翻译得到大规模的中文多模态训练数据
- 结合孟子预训练文本编码模型Mengzi-BERT 初始化
- 参考当前业界领先的OSCAR+框架，将其适用于中文

系统	BLEU
澜舟翻译引擎	
基本Transformer模型 (2千万英汉数据)	39.03
+单语数据 (千万级)	41.65
+单语数据 (亿级)	43.41
+双语平行数据 (亿级)	46.10
+大模型结合知识蒸馏等	48.43
对比翻译引擎	
某著名翻译引擎1	45.97
某著名翻译引擎2	44.79

澜舟翻译引擎性能评测

孟子多模态模型（基于Oscar+）



pre-training input : $(\underbrace{w}_{\text{caption}}, \underbrace{q,v}_{\text{tags\&image}})$ or $(\underbrace{w,q}_{\text{Q\&A}}, \underbrace{v}_{\text{image}})$

pre-training loss : $\mathcal{L}_{\text{Pre-training}} = \mathcal{L}_{\text{MTL}} + \mathcal{L}_{\text{CL3}}$

$\mathcal{L}_{\text{MTL}}$ Masked Token Loss

$\mathcal{L}_{\text{CL3}}$ 3-way contrastive Loss
- the triplet is matched (50%)
- contains a polluted w (25%)
- contains a polluted q (25%)

Xiujun Li, Xi Yin, Chunyuan Li, Pengchuan Zhang, Xiaowei Hu, Lei Zhang, Lijuan Wang, Houdong Hu, Li Dong, Furu Wei, Yejin Choi, Jianfeng Gao. Oscar: Object-Semantics Aligned Pre-training for Vision-Language Tasks. ECCV 2020.

实验数据

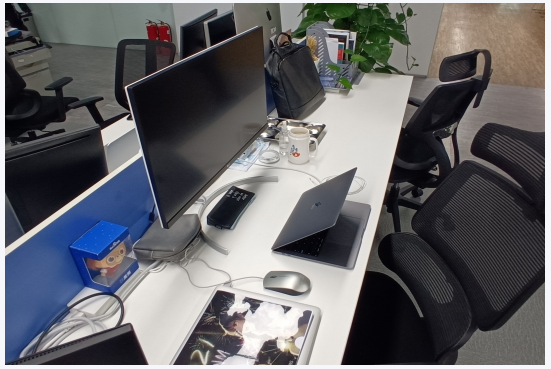
Small	0.22M Images, 2.5M QAs, 0.7M captions							
Medium	1.89M Images, 2.5M QAs, 0.7M captions, 1.67M pseudo-captions							
Large	5.65M Images, 2.5M QAs, 4.68M captions, 1.67M pseudo-captions							
Source	VQA (train)	GQA (bal-train)	VG-QA (train)	COCO (train)	Flicker30k (train)	OpenImages (od train)	CC (train)	SBU (all)
Image/Text	83k/545k	79k/1026k	87k/931k	112k/559k	29k/145k	1.67M/1.67M	3.1M/3.1M	875k/875k
w, q, v	Question, Answer, ImageFeatures			(Generated) Caption, (Generated) ImageTags, ImageFeatures				

轻量化的训练方式：

使用小规模数据 Small corpus：0.22M Images, 2.5M QAs, 0.7M Captions（基于自研机器翻译引擎）

使用了 8 卡，batchsize 设为 1024，训练了 50 万步。

孟子多模态模型（基于Oscar+）



一张白色的桌子，上面放着一台笔记本电脑和一台电脑

绿油油的草地上有两个面带微笑的人在骑马

一条繁忙的城市街道，人们走在人行道上，汽车和摩托车在人行道上

两个打着伞的人和一个背着孩子的男人走在被水淹没的道路上

孟子多模态模型（基于Oscar+）

中文的多模态预训练模型 Mengzi-Oscar

并在下游任务中文图像摘要生成(AIC-ICC)、中文图文互检(COCO-ir)上进行了 fine-tune。

Metrics	BLUE	METEOR	ROUGE-L	CIDEr
UNITER	62.8	38.7	69.2	199.7
文澜	66.1	41.1	71.9	220.7
Mengzi Oscar (Init from Mengzi bert base 0809)	**68.2(+2.1)**	**41.8(+0.7)**	**72.4(+0.5)**	**235.2(+14.5)**

Tasks	I2T Retrieval			T2I Retrieval		
Metrics	R @1	R @5	R @10	R @1	R @5	R @10
CLIP	13.4	27.3	35.1	7.8	18.5	25
文澜	20.3	37.0	45.6	14.4	30.4	39.1
Ours	47.6	70.9	80.8	38.4	65.0	76.2

孟子开源社区微信群

轻量化模型研究展望

数据构造

构造更高价值的训练数据，同等数据下获得**更丰富**的知识

训练目标

探索更好的语言建模策略，让模型**更高效**地获取知识

训练优化

修正训练中的偏差问题，让训练**更准确**，减少无用功或学错知识

结构优化

设计更强大的模型结构，模型**更精简**，减少冗余信息

感谢您的聆听

北京澜舟科技有限公司
www.langboat.com

我们在招聘研究员、工程师、
产品经理和实习生

关注我们