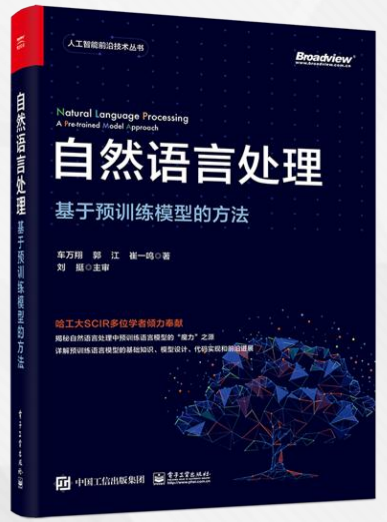


多模态融合的预训练模型

车万翔、郭江、崔一鸣

社会计算与信息检索研究中心
哈尔滨工业大学

目 录

CONTENTS

1 多语言融合

2 多媒体融合

3 异构知识融合

目 录

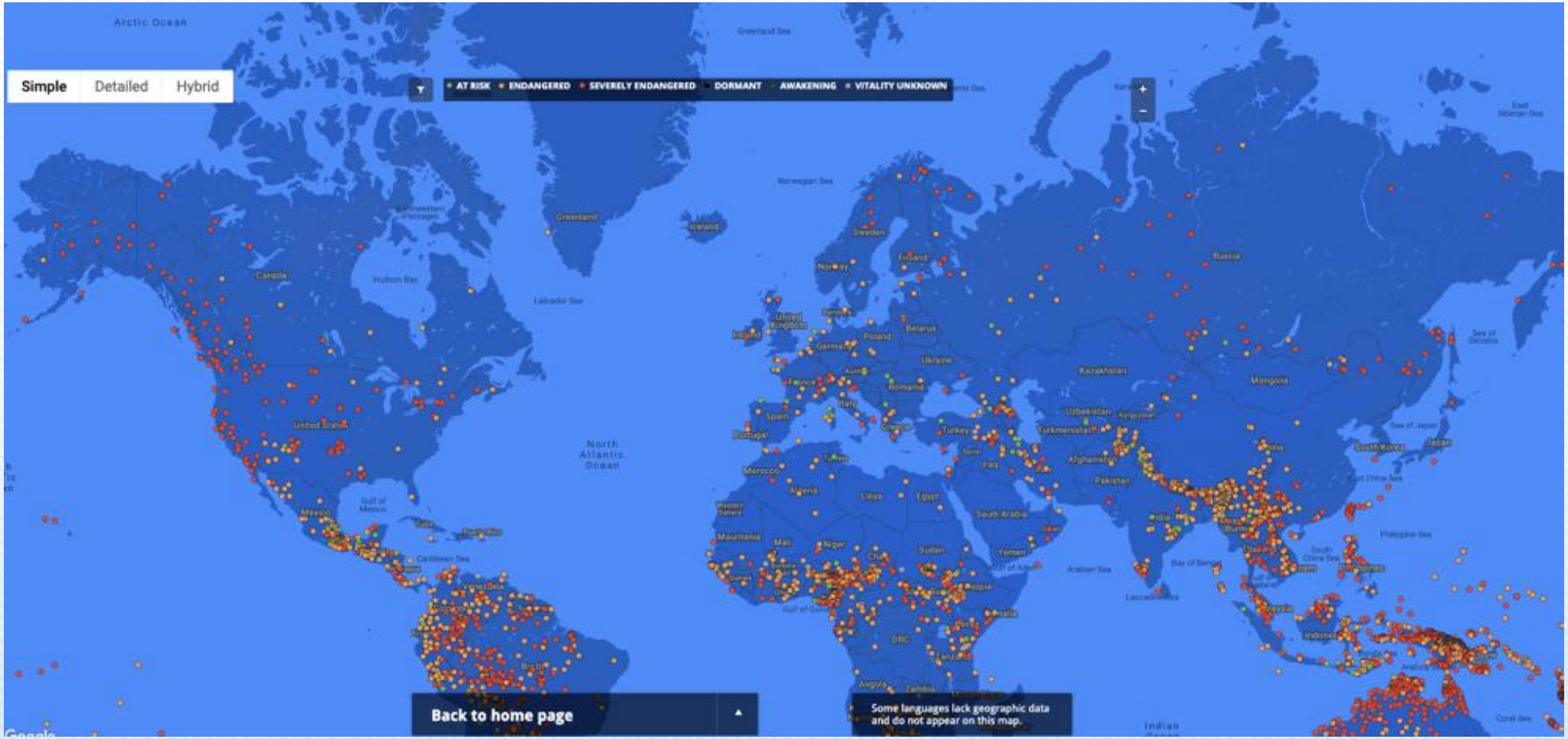
CONTENTS

1 多语言融合

2 多媒体融合

3 异构知识融合

- 超过6,500种语言


Simple
Detailed
Hybrid
AT RISK
ENDANGERED
SEVERELY ENDANGERED
DORMANT
AWAKENING
VITALITY UNKNOWN
Back to home page
Some languages lack geographic data and do not appear on this map.

数据分布的长尾现象（Long Tail Distribution）

Figure credit: Graham Neubig

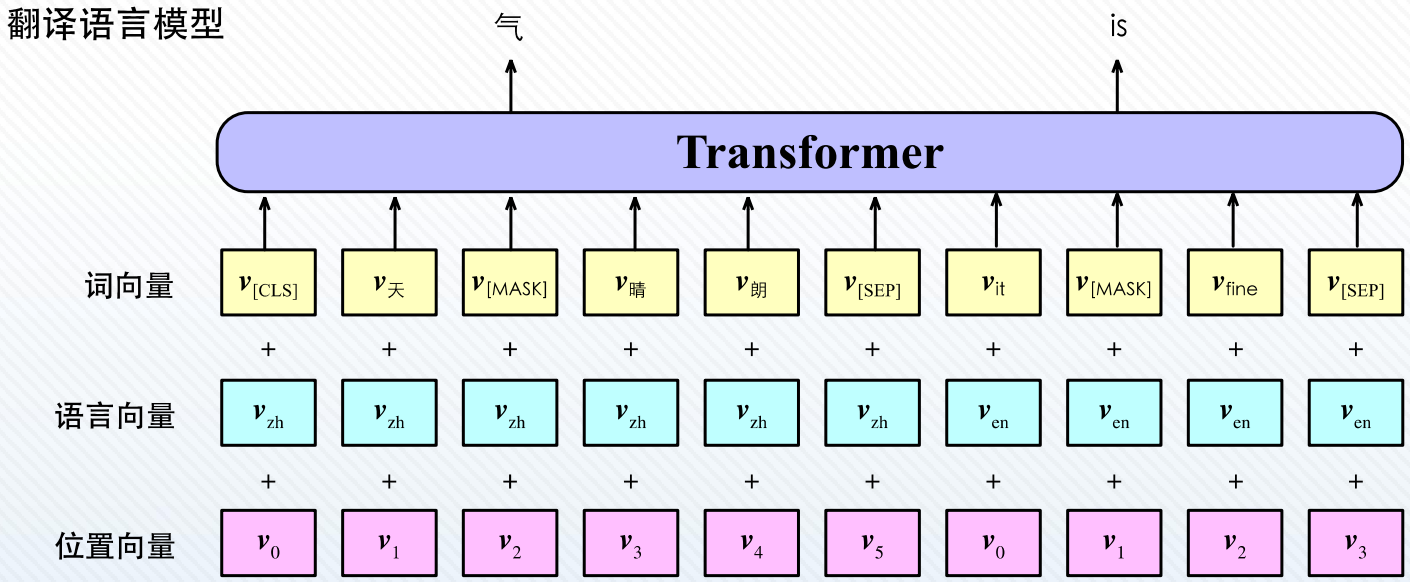
多语言BERT

- Multilingual BERT
  - https://github.com/google-research/bert/blob/master/multilingual.md
  - 统一的多语言表示空间（104种语言）

```
>>> from transformers import pipeline
>>> unmasker = pipeline('fill-mask', model='bert-base-multilingual-cased')
>>> output = unmasker('我like[MASK]')
>>> pprint(output)
[{'sequence': '[CLS] 我 like 你 [SEP]',
  'score': 0.1089084744453430 2,
```

  - 为什么有效?
    - 语言的混合使用（Code-Switch现象）
    - 共享子词

- XLM （Lample and Conneau, NeurIPS, 2019）
  - 翻译语言模型（Translation Language Modeling, TLM）



  - 依赖双语平行句对
    - 大规模平行句对获取难度较高
    - 受限于句子级上下文（篇章/文档级平行数据更为稀少）

- 零样本迁移（Zero-shot Transfer）
  - 将**源语言**（资源丰富，如英语）上训练得到的模型直接应用于**目标语言**（通常为资源稀缺语言）
  - XTREME基准测试集（Hu et al., ICML 2020）

表 9-1　XTREME 数据集的相关信息

任务类型	语料库	数据集规模（训练/开发/测试）	测试集来源	语言数	任务描述
分类	XNLI	392,702/2,490/5,010	翻译	15	文本蕴含
	PAWS-X	49,401/2,000/2,000	翻译	7	复述识别
结构预测	POS	21,253/3,974/47-20,436	独立标注	33	词性标注
	NER	20,000/10,000/1,000-10,000	独立标注	40	命名实体识别
问答	XQuAD	87,599/34,726/1,190	翻译	11	片段抽取
	MLQA	87,599/34,726/4,517-11,590	翻译	7	片段抽取
	TyDiQA-GoldP	3,696/634/323-2,719	独立标注	9	片段抽取
检索	BUCC	–/–/1,896-14,330	–	5	句子检索
	Tatoeba	–/–/1,000	–	33	句子检索

目 录

CONTENTS

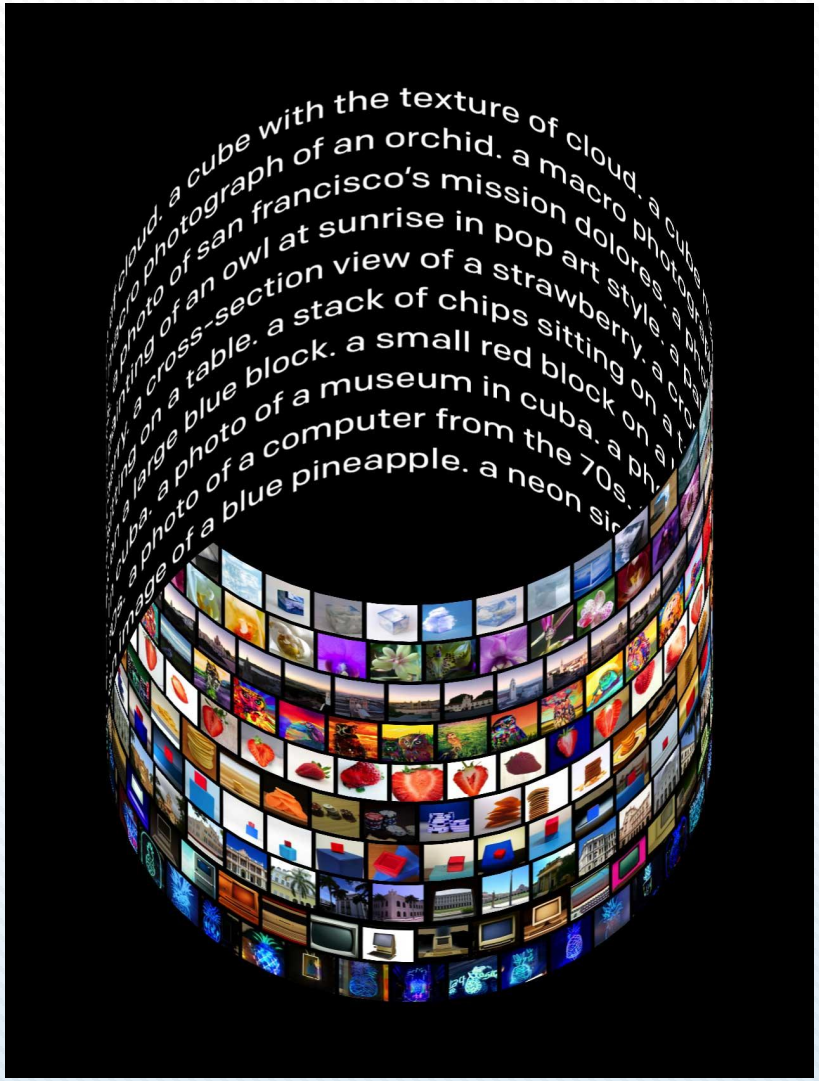
1 多语言融合

2 多媒体融合

3 异构知识融合

- 多媒体数据
  - 语言
  - 图像
  - 视频

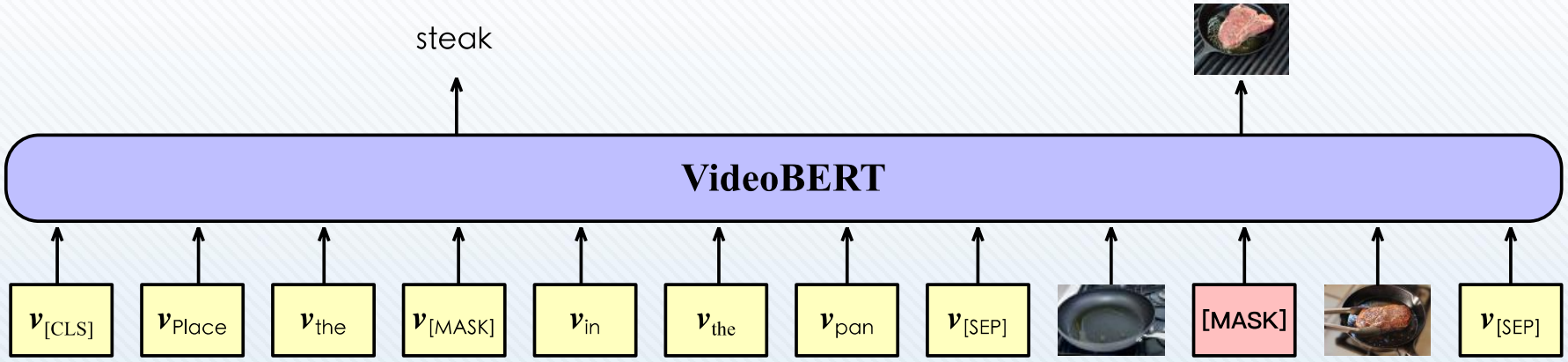
- 跨媒体应用
  - 图像描述生成（Image Captioning）
  - 跨媒体检索（如：以文搜图/视频）
  - 辅助单模态任务
  - ……



图片来源：https://openai.com/blog/dall-e/

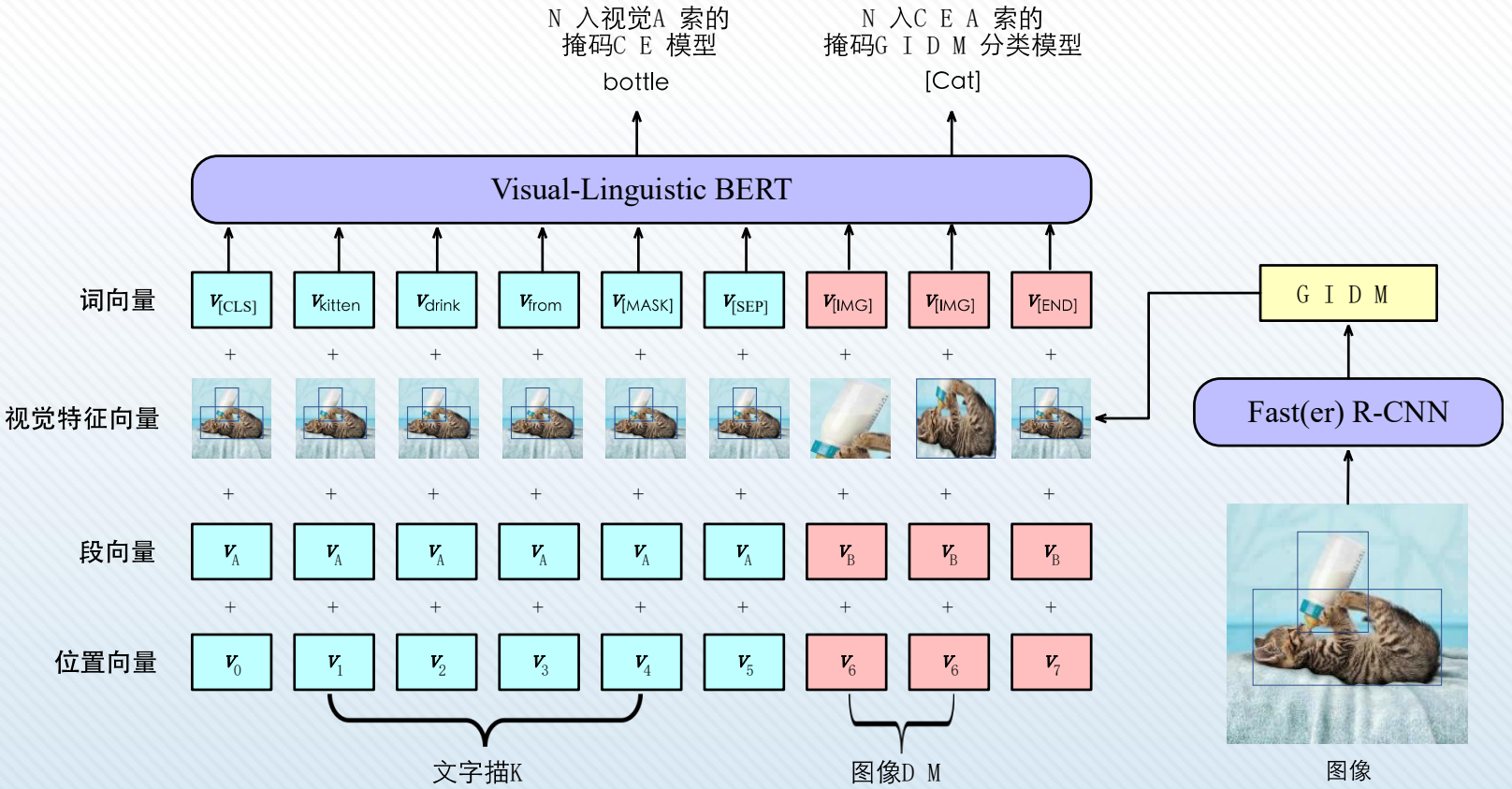
VideoBERT

- Videobert: A joint model for video and language representation learning. (Sun et al., ICCV 2019)
  - “文本+视频”预训练模型
  - 数据：视频以及对应的文本字幕
  - 预训练任务：掩码标记预测



  - 应用
    - 视频检索、视频字幕生成等

- Vl-bert: Pre-training of generic visual-linguistic representations. (Su et al., ICLR 2019)
  - “文本+图像”预训练模型



- Zero-Shot Text-to-Image Generation (Ramesh et al, 2021)
  - 通过离散变分自编码器（Discrete VAE）将图像表示为离散标记序列
  - 由文本生成图像，采用自回归语言模型

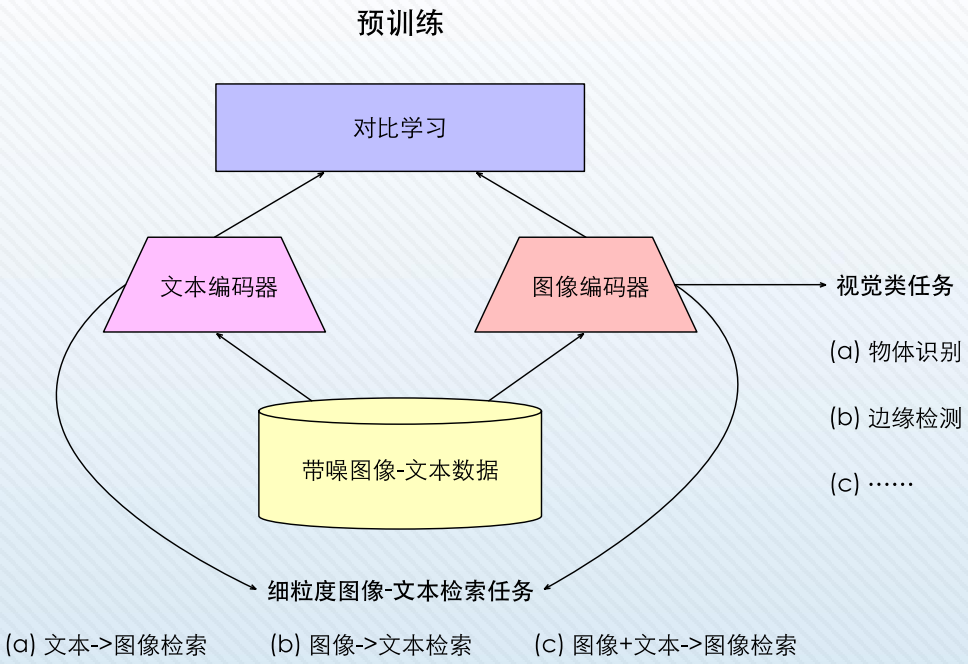
输入：

a clock in the shape of a peacock.
（一个孔形的时钟）

输出：

CLIP，ALIGN

- **CLIP**: Learning Transferable Visual Models From Natural Language Supervision (Radford, et al., 2021)
- **ALIGN**: Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision (Jia et al., 2021)
- 通过对比学习学习图像与文本的联合表示



目 录

CONTENTS

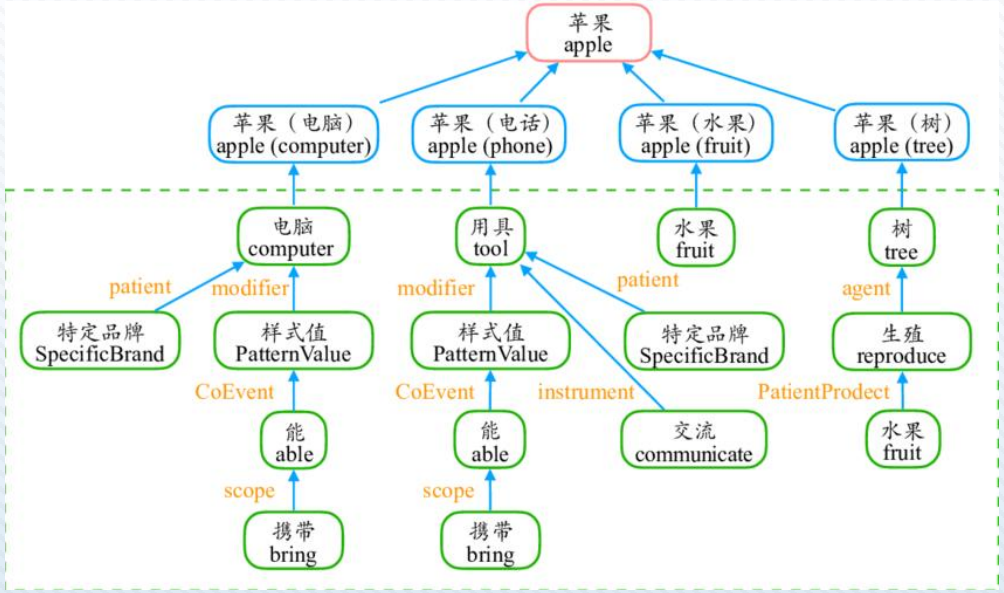
1 多语言融合

2 多媒体融合

3 异构知识融合

知识增强的预训练模型

- “站在巨人的肩膀上”
  - 知识库：（半）结构化知识、外部世界知识、常识知识等
    - 如词典，实体库，知识图谱等
  - 已有任务的标注数据



HowNet

https://zh.wikipedia.org/zh-cn/苹果公司

维基百科

- Ernie: Enhanced representation through knowledge integration (Sun et al., 2019)
  - 预训练任务：掩码语言模型
    - 子词掩码
    - 实体掩码
    - 短语掩码

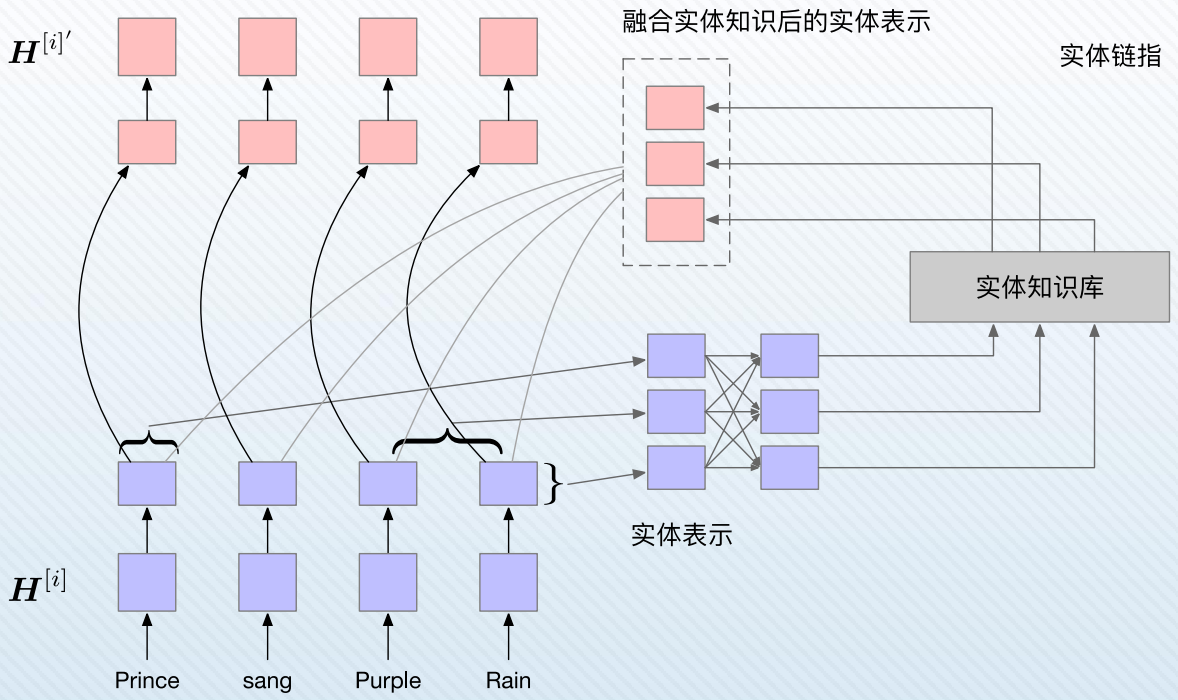
- 示例：

原始句子	Harry Potter is a series of fantasy novels written by J. K. Rowling
子词级别掩码	[M] Potter is a series [M] fantasy novels [M] by J. [M] Rowling
实体级别掩码	Harry Potter is a series of fantasy novels written by [M] [M] [M]
短语级别掩码	Harry Potter is [M] [M] [M] fantasy novels [M] by [M] [M] [M]

- 有助于实体间关系的学习，如上例（Harry Potter，J. K. Rowling）

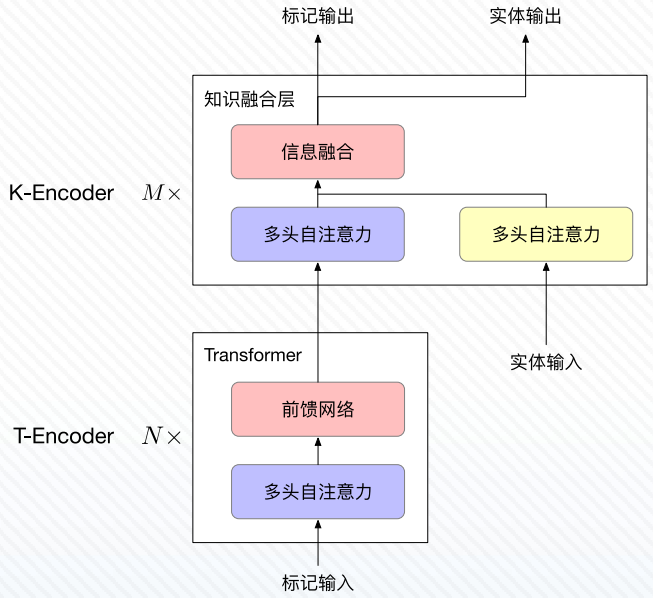
命名实体

- KnowBERT: Knowledge enhanced contextual word representations (Peters et al., EMNLP 2019)
  - 利用实体消歧（Entity Linking）以及相应的文本描述
  - 在BERT相邻两层Transformer之间引入知识融合模块

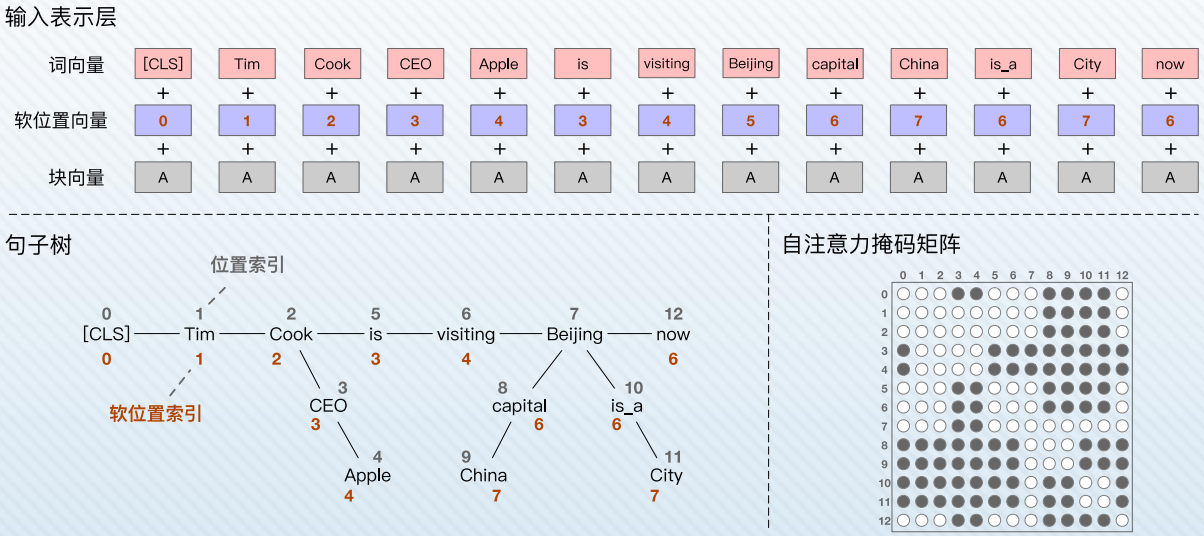


知识图谱

- ERNIE^{THU} (Zhang et al., 2019)
  - 知识编码器（K-Encoder）
    - 融合文本与知识图谱中的实体表示
    - 利用TransE获取知识图谱中的实体表示



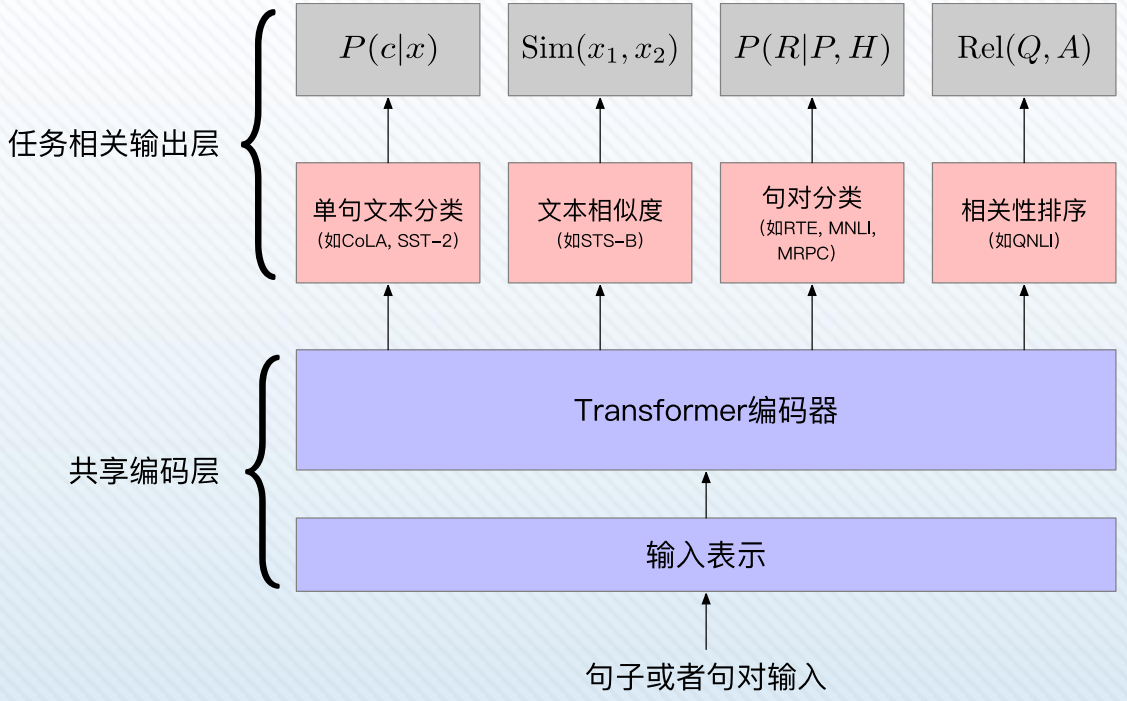
- K-BERT
  - 推理阶段的知识融合
  - 树状结构输入（根据知识图谱对实体进行扩展）
  - 对自注意力分布进行约束



多任务知识

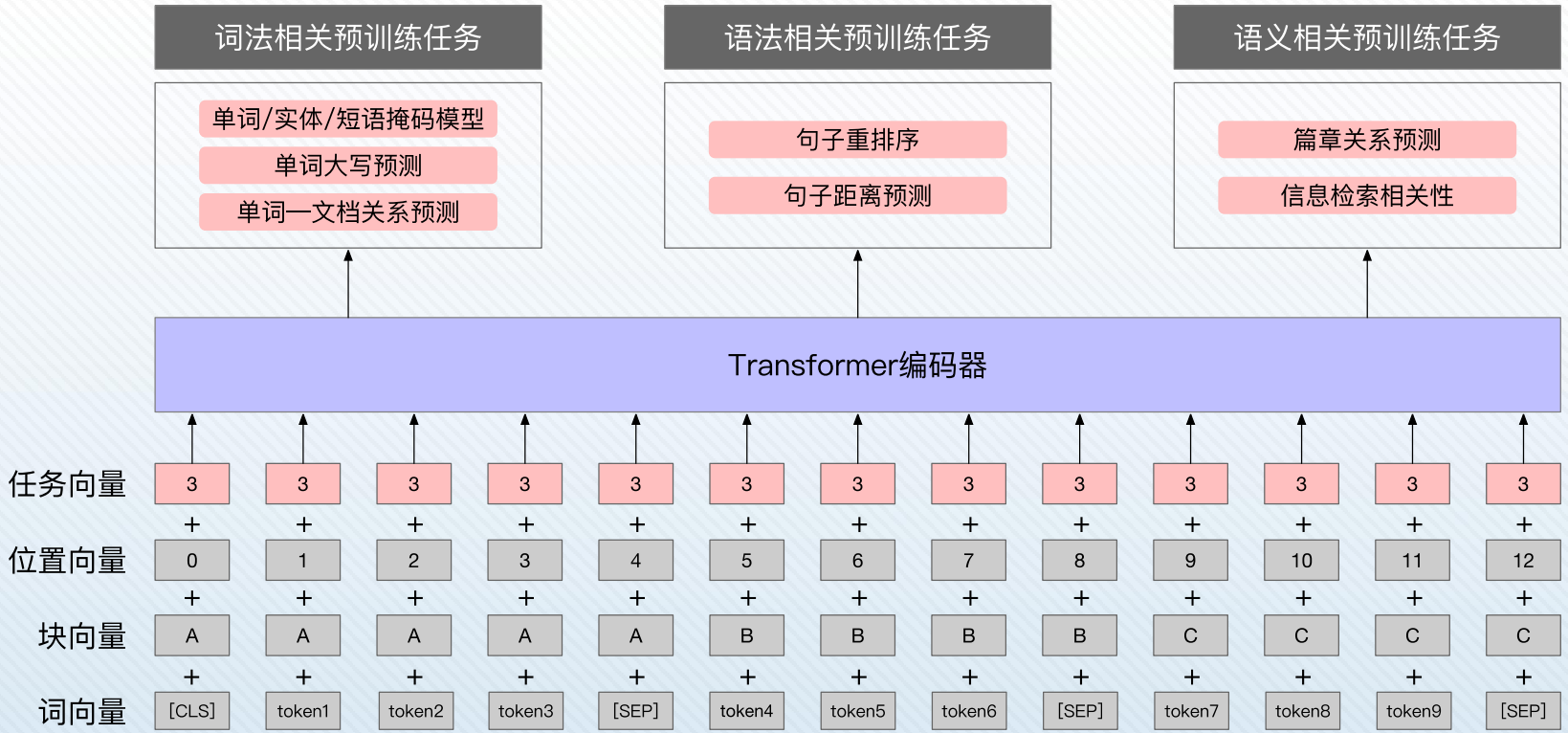
- 充分利用已有任务的标注数据及资源
  - 文本分类：如情感分类
  - 回归问题：句子相似度预测
  - 句对分类：文本蕴含
  - 与基于海量无标注数据的自监督预训练相媲美

- MT-DNN：Multi-task Deep Neural Network (Liu et al., ACL 2019)
  - CoLA, SST-2
  - RTE, MNLI, MRPC
  - QNLI



更丰富的预训练任务

- ERNIE 2.0 (Sun et al., AAAI 2020)
  - 词法，语法，语义层面分别设计预训练任务
  - 连续多任务学习（Continual Multi-task Learning）



谢 谢!

理解语言，认知社会

以中文技术，助民族复兴

长按二维码，关注哈工大SCIR

微信号：HIT_SCIR