

Python程序设计与数据科学导论

—— 课程内容回顾与总结



胡俊峰 北京大学

2023/06/08



内容

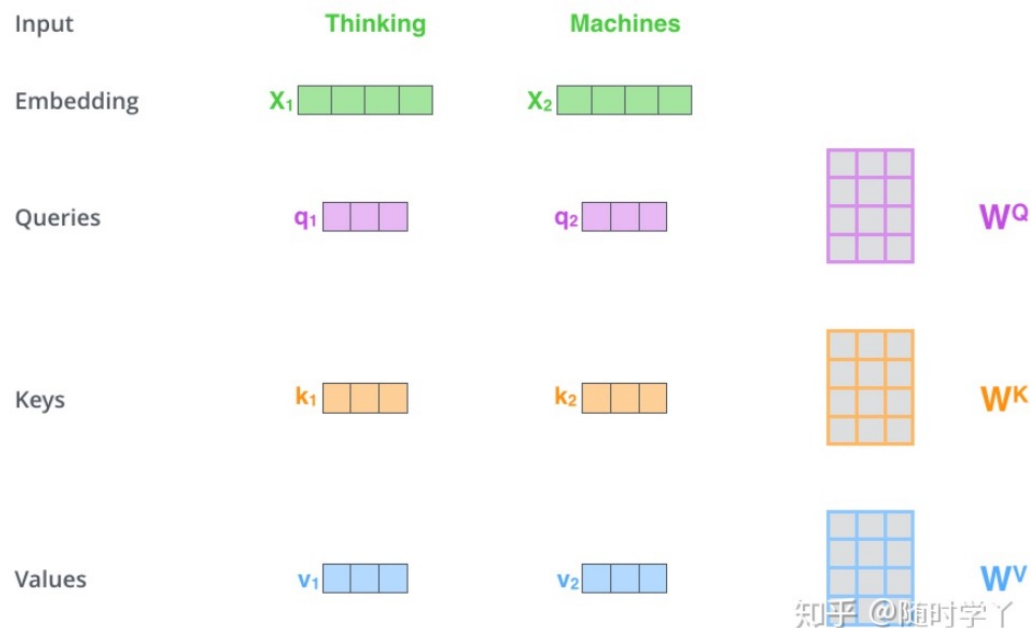
- Embedding is language model (小结次课内容)
- 课程内容回顾
- 考试安排与成绩评定

根据上下文特征的 Embedding → NLU

Step 1

首先给出信息输入：用 $X = [x_1, x_2, \dots, x_n]$ 表示 N 个输入信息，通过线性变换得到 Q, K, V 三个向量的初始表示：

$$\begin{aligned} Q &= W_q X \\ K &= W_k X \\ V &= W_v X \end{aligned}$$



<https://zhuanlan.zhihu.com/p/109983672>
<https://www.zhihu.com/question/350369171>

transformer中multi-head attention中每个head为什么要进行..



海晨威
微软 算法工程师

+ 关注

217 人赞同了该回答

一言蔽之的话，大概是：在**不增加时间复杂度**的情况下，同时，借鉴**CNN^Q多核**的思想，在**更低的维度**，在**多个独立的特征空间**，**更容易**学习到更丰富的特征信息。

前面多位答主也都提到了，在经过维度的“分割”之后，在多个低维空间^Q，相比原有的高维空间，能降低特征学习的难度。这里再补充一下**时间复杂度**的计算：

Self-Attention时间复杂度： $O(n^2 \cdot d)$ ，这里， n 是序列的长度， d 是embedding的维度。

Self-Attention包括三个步骤：**相似度计算^Q**，**softmax**和**加权平均**：

$$\text{softmax} \left(\frac{Q \times K^T}{\sqrt{d_k}} \right) V$$

Resulting in matrix Z :

$$= Z$$

Embedding的最优词语序列生成 → NLG

	GPT-4	ChatGPT / GPT-3.5
发布时间	2023年3月14日	2022年11月30日 (ChatGPT)
模态对比	输入: 图像+文字 输出: 文字	输入: 文字 输出: 文字
文字逻辑推理能力	■■■■■	■■■
应答速度	■■	■■■■■
应答简洁度	■■■■	■■
图文分析或解答	有	无
模拟律师考试分数	前10%	末10%
最新数据日期	2021年9月	2021年9月
模型参数	未知	大概175B-6B-1.3B
上下文窗口	32,000 token (相当于24000单词)	4096 token (相当于3072单词)
主要算法技术	1) PPO算法微调 2) RLHF (人类反馈强化学习, 使用奖励模型) 基于规则的奖励模型 (RBRM)	PPO算法微调 RLHF (人类反馈强化学习, 使用奖励模型)
基本架构 (猜测)	基于Transformer结构, 具备局部带状稀疏注意模式 视觉语言模型组件 (VLM)	基于Transformer结构, 具备局部带状稀疏注意模式
训练方法	1) 有监督微调 2) RLHF训练奖励模型 3) 构造基于规则的奖励模型 (RBRM) 4) PPO强化学习	1) 有监督微调 2) RLHF训练奖励模型 3) PPO强化学习

陈巍谈芯

Hinton关于人工智能的担忧

2.2 GPT-4的多模态架构

2.2.1 多模态对于大语言模型的重要意义

2.2.2 GPT-4多模态架构与交叉注意力

2.3 GPT-4的关键技术

2.3.1 理论基础——多模态涌现能力

2.3.2 核心优势——多模态思维链

2.3.3 编程范式——多模态提示工程

2.3.4 关键技术——人类反馈强化学习

2.3.5 安全技术——基于规则的奖励模型

2.3.6 安全技术——多模态幻觉检测

2.3.6 安全技术——多模态幻觉检测

大型语言模型（Large Language Model, LLM）的**幻觉**（Hallucination）指的是模型生成的输出包含一些与输入不符合的信息，这些信息可能是错误的、无关的或者荒谬的。与人类直觉相反，随着模型变得更加以假乱真，幻觉会变得更加危险。GPT-4等模型的这种幻觉可能会出现在各种类型的任务中，比如文本生成、图文分析和问答系统等。

由于大模型（包括GPT-4）本质上可以视为训练集（人类知识/语言）的**有损压缩**，因此在模型运行时无法完整复现或者应答原始知识，从而模型的幻觉来自于信息压缩的偏差。多模态幻觉的本质是这种有损压缩偏差的体现，也是通过**数学逼近人类语言**的必然代价。（类似于压缩后的图像边缘出现不正常的条纹）模型场景的增加和模型参数数量的巨大也增加了幻觉出现的概率。

Frank Herbert's Dune: "Thou shalt not make a machine in the likeness of a human mind."



课程内容总结与回顾：课程总览

- Python语言编程
- Pandas数据处理与分析软件包
- Numpy、矩阵运算、线性空间模型、卷积运算
- 图像特征提取
- 基于观影数据的分析与推断
- 时间序列分析与序列模型回归
- Pytorch深度学习框架；CNN、RNN基础、网络模型参数优化与压缩
- Word embedding-attention-NLP综述



总体评价：

- 课程主要内容顺利完成
- 新搭建的大作业平台初步成型
- 作业提交、作业批改、讲评方案有待改进

具体内容回顾：Python语言部分

- 容器-迭代器、生成器 ★ ★ ★
- 函数、匿名函数与列表生成式、函数式编程 ★ ★ ★
- 类定义、实例化、继承 ★ ★
- 可执行类、迭代器类、装饰器类 ★ ★
- 进程、线程 ★
- 协程 ★ ★ ★
- 设计模式 ★
- 网络编程 ★



Pandas数据处理

- 表数据处理基础、apply函数使用、缺失数据处理 ★ ★ ★
- 聚合操作与统计、表连接-融合 ★ ★ ★
- 高阶数据处理技术：高维表变换、透视表处理 ★ +



Numpy

- 矩阵基本操作：定义、维度信息、维度交换、矩阵切片 ★ ★ ★
- 矩阵运算、广播机制、向量运算 ★ ★ ★
- 矩阵操作内置函数 (ufun) 使用 ★ ★
- 矩阵线性空间运算 ★ ★



Python与机器学习

- 贝叶斯分类器 ★ ★ ★
- 线性回归、K-means、KNN ★ ★
- 图分析基础 ★
- 序列回归（线性）模型 ★
- 神经网络模型实现 ★ ★
- 深度学习架构实现 ★
- Pytorch张量计算★、CNN、RNN ★ ★
- 深度学习建模概念 ★ ★

基于观影数据的类型化分析与预测

- 性别信息
- 职业信息
- 观影-评分信息
- 电影类型信息
- 电影摘要信息
- 电影年代-导演-演员
- 电影海报信息



课程内容回顾（划重点）

- Python语言技术（30-35%）
- Numpy、pandas及机器学习相关概念（25-30%）
- CV、NLP、SNA、图分析（10%-15%）
- 深度学习及相关矩阵运算建模（15-20%）
- 期中大作业相关：10%



课程总结：

- 大作业内容有所调整，有些话题没来得及展开。

■ 期末考试时间-地点：

- 考试时间为：2023/06/12，星期一
- 考试时间：下午，14:00 – 16:00
- 考试地点：二教101
- 考试内容：主要是课上讲的基础内容。重点考察作业涉及的知识和技术。
- 考试形式：闭卷笔试

关于缓考

- 学生缓考要在提交缓考申请单，**一律由学院审批；**
学院批准的缓考，会要求学生把缓考回执单交给任课教师。
- 课程开考后递交的缓考申请和病假条无效。
- 任课教师在考场记录上标注该生缓考，成绩系统中的缓考由教务老师录入。



感谢助教团队的辛勤工作！

感谢同学的积极参与与合作！

预祝大家取得好成绩！

