

戴雄杰

+1-217-305-0416 | @xdai12@illinois.edu | linkedin.com/in/xiongjie-dai | github.com/Xiongjiedai

教育背景

伊利诺伊大学厄巴纳-香槟分校

理学硕士，统计学；绩点：3.84/4.00

伊利诺伊州, 美国

2021 年 8 月 - 2023 年 5 月

暨南大学 / 伯明翰大学

经济学学士；绩点：3.93/4.25

广州, 中国/伯明翰, 英国

2017 年 9 月 - 2021 年 6 月

应用数学与统计学理学学士（一等学位）

技能

机器学习研究：人工智能、深度学习、自然语言处理、统计学习、计算机视觉

数据科学：数据科学编程方法、统计数据管理、统计咨询、统计模型

编程：Python、R、SQL

技术工具：Git、AWS、Shell、PyTorch、TensorFlow、NumPy、Pandas、Matplotlib、ggplot2、Tidyverse、Gradio、Docker

工作经历

人工智能/机器学习工程师

Git, Shell, RunPod, LLM, GenAI

Wisping LLC, 伊利诺伊州

2023 年 7 月 - 现在

- 设计并实施全面的 LLaMA 模型测试，比较 NVIDIA GPU 和 Apple Silicon 的性能，采用云计算平台评估性能，为设备上的大型语言模型（LLM）推理建立概念验证；项目获得开源界认可，GitHub 上获得超过 **600** 颗星标。
- 记录和分析不同模型大小在各种量化和多种硬件配置下的推理速度，提供关于高效 LLM 推理的最佳设置的可操作见解。
- 协助 AI 爱好者从零开始入门本地大模型推理，同时对其它开发者提出的疑问进行解答。

研究生课程助教

Python, R, SQL, 深度学习, 统计学习

伊利诺伊大学厄巴纳-香槟分校

2022 年 9 月 - 2022 年 12 月

- 使用 Python, R 和 SQL 与教授和其他助教合作设计富有吸引力的机器学习课程内容和创新的编程能力评估作业；在每周与学生的 office hour 期间批改作业，给学生的作业提供反馈，并回答学术上的问题，期间课程零投诉。

项目

神经网络机器翻译：RNN 和 Transformer 模型 | Python, PyTorch

2023 年 5 月

- 领导开发神经网络机器翻译（NMT）系统，使用 PyTorch 设计 RNN 和 Transformer 模型，专注于西班牙语-英语的翻译任务。
- 设计并实施了采用 GRU 和注意力机制的 RNN，构建编码器和解码器组件。在 Transformer 模型中通过多头注意力和位置编码增强翻译质量，以捕捉句子结构并保持词序。
- RNN 模型和 Transformer 模型分别达到了 0.058 和 0.059 的 BLEU-4 分数。

强化学习在贪吃蛇游戏上的应用 | Python, Numpy

2022 年 12 月

- 开发了一个强化学习 AI 代理，采用 TD Q-learning 算法，在明确的马尔可夫决策过程（MDP）框架中逐渐掌握贪吃蛇游戏。
- 实施探索政策算法，平衡探索新状态和利用已知状态，确保代理在测试环境中全面的学习覆盖，并使用递减的学习率优化随时间的 Q 值更新过程。
- 通过严格的测试阶段分析代理的性能，确保模型在不同测试环境下的稳健性和适应性。

亚马逊评论和药品数据集的情感分析 | R markdown, Word2Vec

2022 年 12 月

- 对亚马逊和药品评论数据集进行高级情感分析，比较四种经典嵌入和 NLP 方法（BoW、Word2Vec、GloVe、fastText）。讨论了 FastText 在处理未见词时显示出优越性（在亚马逊数据集上的准确率为 86.49%，在药品评论数据集上的准确率为 78.69%）。
- 采用数据预处理技术进行文本标准化和向量化。使用朴素贝叶斯和随机森林算法对文本进行分类。

探索数据增强技术对不同 CNN 模型的影响 | Python, PyTorch, Matplotlib, Gradio

2022 年 5 月

- 开发并评估三种不同大小的卷积神经网络（CNN）模型，以确定数据增强技术对模型性能的影响，模型专注于猫狗图片二分类挑战。
- 模型应用正则化方法（dropout, L1, L2）避免过拟合，以二元交叉熵为成本函数，SGD 为优化器；通过图片翻转技术对数据增强，最佳 CNN 模型的准确率提高约 15%。