

唐小宇

tangxiaoxyu_work@outlook.com

18365616569

人工智能方向计算机博士毕业。熟悉数据分析，机器学习，深度学习等领域。熟练使用多种编程语言、云计算平台以及开源大数据工具如 Spark, Hadoop 等。能够快速学习掌握新知识并付诸实践。具有团队合作精神，创新意识，有进取心，对工作认真负责，可适应高强度的工作。

专业技能

✓ 机器学习、深度学习、统计学:

- 自然语言处理 (NLP) / 大语言模型 (LLMs): word2vec, Attention mechanisms, Transformers, BERT, GPT, LLaMA, BLOOM, LLM 微调, instruction tuning, PEFT, Adapter Tuning, LoRA, QLoRA, Prefix Tuning, P-Tuning, Prompt Tuning, Retrieval Augmented Generation (RAG), Hybrid RAG, Adaptive RAG, Graph RAG, Agentic RAG, AI Agentic Workflows, ReAct, ReWoo, etc.
- 推荐系统 (Recommendation Systems): MF, LFM, TF, FM, Wide&Deep, DeepFM, ESMM, DBMTL, Learning To Rank (LTR), etc.
- 深度学习 (Deep Learning): CNN, RNN, LSTM, GRU, Autoencoders, Deep Embedded Clustering
- 分类算法 (Classification): Decision Trees, Logistic Regression, SVM, KNN, Naive Bayes, Neural Networks
- 集成学习 (Ensemble Learning): Random Forest, AdaBoost, GBDT, XGBoost, LightGBM
- Dimensionality Reduction: Matrix Decomposition, Tensor Decomposition

✓ ML Libraries:

- PyTorch, TensorFlow, HuggingFace, LangChain, scikit-learn, numpy, pandas, matplotlib, seaborn, Jupyter Notebook

✓ MLOps:

- Docker, Argo Workflows, GitHub Actions, MLflow, WandB, TensorBoard, etc.

✓ 云计算、大数据平台:

- Amazon Web Services (AWS), Google Cloud Platform (GCP), Spark, Hadoop, Hive, Zeppelin, Hue

✓ 编程语言、数据库等:

- Python, Scala, Java, C#, SQL, DynamoDB; 以及 R, Matlab, MongoDB 等。

工作经历

个人项目

2023/08 – 现在

使用 DeepSeek 及 Llama 等大模型实现大语言模型的相关应用，如 Hybrid RAG, Adaptive RAG, Agentic RAG, agentic workflows 等等

Data Scientist (算法工程师)

2020/11 – 2023/07

Woolworths Group | 澳大利亚悉尼

公司简介：Woolworths Group 是一个拥有多个知名零售品牌的公司，主要经营澳大利亚的 Woolworths 超市、新西兰的 Woolworths (以前称为 Countdown) 以及百货商店 Big W 等。WooliesX 是 Woolworths Group 的数字化部门。

主要用到的技术包括: GCP, Python, SQL, Kubernetes, Argo Workflows, PyTorch, HuggingFace Transformers, BERT, LightGBM, Google OR-Tools, 等。

参加了 Big W (大型连锁折扣超市) 推荐系统管道架构的实现和运行维护，其中主要使用了 Kubernetes, Argo Workflows, LightGBM 和 GCP 等技术。

结合 NLP 技术设计实现了神经网络模型（dual encoder interest network）用于商品的个性化推荐，其中主要用到的技术包括预训练语言模型 BERT 及微调，Transformers，PyTorch，HuggingFace 等。通过 DDP 方式使用多 GPU 进行分布式数据并行模型训练。

参加了 Smart Substitution 项目，主要使用了 PyTorch, BERT, Huggingface Transformers 等技术。该项目的目标是为在线用户购买的缺货商品自动化推荐可替代商品，以降低人工费用。

参加了 Woolworths eStores 的货位优化管理项目（slotting optimization）。该项目的目标是，通过优化商品的存储区域来提高订单拣选的效率。根据业务需求实现了多种货位优化算法，提高了自动订单拣选系统和其他高效仓库存储系统的利用效率。项目所使用到的技术主要包括 Python, pandas, matplotlib, Google OR-Tools 等。

Data Scientist (算法工程师)

2017/01 – 2020/10

Symplicity (CareerHub) | 澳大利亚布里斯班

公司简介：Symplicity 是 CareerHub 背后的品牌，它是全球领先的学生和毕业生招聘平台之一。Symplicity 为 37 所澳大利亚大学的学生和校友提供了与雇主连接的机会。CareerHub 是一个集中式平台，允许招聘人员在 Symplicity 广泛而多样化的大学网络上发布工作。

主要用到的技术包括：AWS, Python, C#, ASP.NET Web API, SQL, Scala, Spark, TensorFlow, NLTK, Word2Vec, TF-IDF，等。

基于多任务学习和深度学习，为 CareerHub 设计实现了一个职位信息推荐系统，用于 CTR 和 CVR 预测任务。模型架构主要参考基于 Entire Space Multi-task Model (ESMM) 和 Deep Factorization Machines (DeepFM) 进行设计实现。部分特征的 embeddings 在两个任务目标之间进行共享。模型实现主要用到的框架和工具是 TensorFlow 2 和 Amazon SageMaker，数据预处理流程主要用到的工具和平台是 Spark (Scala), Amazon EMR 等。

使用 C#, SQL server 实现了一个协同过滤推荐系统。为了解决冷启动问题，使用 Spark (Scala) 实现了一个分布式的基于内容的推荐系统，并通过 Amazon EMR, Lambda, DynamoDB 把推荐系统部署在 AWS 云计算平台。

使用 NLTK, TensorFlow 等实现了 word2vec 模型用于对发布职位的文本信息进行建模。

项目研究经历

基于最近邻和张量分解的多维推荐系统框架

2012 – 2016

博士研究生 | 昆士兰科技大学 (*Queensland University of Technology*) | 澳大利亚

- ✓ 基于矩阵分解，张量分解技术及核方法等设计实现了多种新算法，提高了协同过滤推荐预测精确度。
- ✓ 使用 Matlab 和 Matlab Tensor Toolbox，实现了高阶张量建模算法。
- ✓ 使用 Python, NumPy 和 SciPy 等，基于矩阵分解 SVD 的相关性质实现了基于张量分解的推荐算法中对用户兴趣建模进行增量更新的计算方法。
- ✓ 设计了基于贝叶斯概率的用户偏好模型和推荐算法，使用 Java 以及开源框架工具（Mahout, Guava, JUnit, Maven 等）实现了相关算法，并使用社会化标签系统（social tagging systems）的多种数据集在 High Performance Computing (HPC) 上进行了实验验证。
- ✓ 使用 R, ggplot2 等进行了统计分析和数据可视化。
- ✓ 完成相关技术报告，并在国际期刊和会议中发表多篇论文。

用户兴趣模型驱动的科技论文在线推荐系统（项目编号 Grant No. 20093718110008），教育部网络时代的科技论文快速共享研究项目

2010 – 2010

硕士研究生 | 山东科技大学

本人作为项目组组长，主要进行了以下工作：

- ✓ 基于 HCS 聚类设计开发了一个关键词网络聚类的自动生成算法；改善了传统的基于本体的推荐算

法，并在用户兴趣模型中使用了显式用户模型和隐式用户模型的组合，提高了基于内容的个性化兴趣推荐的准确性与意外发现特性(serendipity)

- ✓ 设计开发了一个科技论文共享与推荐系统，并使用 C#, ASP.NET 和 SQL server 实现了系统主要功能
- ✓ 组织并指导了项目的代码实现

基于 Petri 网的并行程序行为验证（项目编号 No. ICT-ARCH200807），

2009 – 2010

中科院计算机体系结构实验室开放基金

硕士研究生 | 山东科技大学

本人参与了部分研究，主要工作内容为：

- ✓ 基于 Ubuntu Linux 操作系统下的 MPICH2 建立了一个计算机集群系统
- ✓ 用 C++建立了一组的 MPI 程序，并将其翻译成 Promela 模型
- ✓ 使用 MPI-Spin 1.1 验证 MPI 程序，并为项目组提供了测试报告

教育经历

计算机科学博士（Doctor of Philosophy in Computer Science）

昆士兰科技大学（Queensland University of Technology） | 澳大利亚

2012 – 2016

计算机软件与理论专业硕士

山东科技大学

2008 – 2011

网络工程专业学士

山东科技大学

2004 – 2008

曾获奖励

优秀硕士学位论文，山东省教育厅

2012

研究生优秀科技创新成果奖，山东科技大学

2011

2010 年美国大学生数学建模竞赛二等奖

2010