应佳欣

https://jiaxinying.me • +1 (734) 881-5130 • jiaxinyi@umich.edu

教育背景

- **密西根大学** 美国，密西根，安娜堡
  数学学士 & 计算机科学学士；GPA：3.98/4.0 *2019.09 – 2021.05*
  课程：算法，数据库系统，机器学习，数据挖掘，网络系统，搜索引擎，概率论，线性代数，随机过程，图论
- **上海交通大学** 中国，上海
  电气与计算机工程学士；GPA：3.8/4.0 *2017.09 – 2021.08*
  课程：高等数学，程序设计与数据结构，数值分析，工程概率方法

科研经历

- **模糊特征图神经网络及其在图级监督学习中的应用** 美国，密西根，安娜堡
  密西根大学 Foreseer 实验室研究助理；导师：Qiaozhu Mei 教授 *2019.10 – 2020.12*
  - 研究概述：提出了一个新型的，基于模糊特征的图神经网络（FF-GNN），通过多通道模糊直方图将图参数化，使之与传统的图特征方法及 MP-GNN 相比具有在预测准确率、时间复杂度、GPU 内存使用上更好的综合性能。
  - 使用 PyTorch 实现了参数可学习的模糊直方图，将节点级特征转化成全局图级表示方式。
  - 运用 TensorBoard 和 NetworkX 将神经网络隐藏层和图的结构可视化，以找出模型性能的限制点。
  - 设计了在人造数据集和现实场景数据集上的测试实验，通过实验室的 CUDA 环境和 AWS 的 GPU 实例运行实验，将 FF-GNN 和其他基准模型的各项指标进行比较。
  - 撰写了论文的引言、方法与实验部分的初稿，并在实验室组会中发表论文与研究工作的总结报告。
- **图像定位词汇学习中的物理因果关系** 美国，密西根，安娜堡
  密西根大学情境语言与具象对话（SLED）实验室研究助理；导师：Joyce Y. Chai 教授 *2020.05 – 至今*
  - 研究概述：受到关于幼儿语言学习过程的认知科学启发，发现了对动作过程中物理因果关系的理解能够为动词和名词的学习效果带来 5.2%−13.3% 的交叉提升。
  - 对 ACT 数据集进行边界框和注释对应关系的人工标注，以及数据清理工作。
  - 参考 Faster R-CNN 和注意力机制，设计并实现了两种图像定位模块，通过输入基于 ResNet、BERT、word2vec 的图像及文字嵌套，输出对于边界框的预测。
  - 撰写了论文的相关工作部分，并提出了对其他部分的多条建议。

项目经历

- **推特社交网络数据集的 Spark 迭代处理** *2020.12*
  - 通过 Zeppelin Notebook，使用 Spark RDD 和 DataFrame 对推特社交网络数据集进行分析，并实现了 PageRank 算法计算出该网络中影响力最大的用户。
  - 运用 YARN 和 Spark UI 对程序进行调试，例如 DAG 可视化以找出性能瓶颈。
  - 将 Spark 应用部署到 Azure Databricks，使之相比于 HDInsight 节省了 12% 的运行时间。
- **EECS 484 数据库管理系统课程项目** *2020.01 – 2020.04*
  - 针对类 Facebook 应用设计关系型数据库的表结构，并使用 Java 和 JDBC 开发了多个 SQL 查询功能，包括用户信息检索、附近活动与新朋友推荐。
  - 将该 Facebook 数据集从 Oracle SQL*Plus 迁移到 MongoDB，以支持 JSON 形式的更灵活的数据库结构，同时用 JavaScript 重写了查询语句。
- **STATS 415 数据挖掘课程团队项目：乳腺癌数据集分析** *2019.11 – 2019.12*
  - 基于威斯康星州乳腺癌数据集，分析细胞核特征数据，应用主成分分析（PCA）探究其中最关键的预测变量组合。
  - 用 R 语言实现了逻辑回归、随机森林、SVM、LDA、QDA 等模型并比较其在数据集上的性能，得出 QDA 结合最优子集选择能够对病例诊断进行最有效的预测。
- **VE 280/EECS 281 数据结构与算法课程项目** *2019.05 – 2019.12*
  - 利用类的继承和动态多态性质，实现了基于排序数组、二叉堆、配对堆的优先队列，并支持模板与泛型。
  - 应用分支限界法和最小生成树（MST）解决完全带权图中的旅行商问题（TSP）。
  - 编写了 C++ 版本的 2048 小游戏，使用户通过键盘按键进行操作，同时支持表情符号等自定义滑块类型。

专业技能

- **编程语言**: Python, C++, R, Scala, Java, SQL, MATLAB
- **其他技能**: PyTorch, TensorFlow, scikit-learn, MySQL, SQL*Plus, MongoDB, Spark, AWS, Azure, Linux, Docker, Git

JIAXIN(RACHEL) YING

https://jiaxinying.me • +1 (734) 881-5130 • jiaxinyi@umich.edu • MI 48105

EDUCATION

- **University of Michigan** Ann Arbor, MI
  B.S. Mathematics & B.S. Computer Science; GPA: 3.98/4.0 *September 2019 – May 2021*
  Courses: Algorithms, Database Systems, Machine Learning, Web Systems, Search Engines, Probability Theory, Linear Algebra
- **Shanghai Jiao Tong University** Shanghai, China
  B.S.E. Electrical and Computer Engineering; GPA: 3.8/4.0 *September 2017 – August 2021*
  Courses: Honors Mathematics, Programming & Elem. Data Structures, Numerical Methods, Probabilistic Methods in Eng.

RESEARCH EXPERIENCE

- **Fuzzy Fingerprinting Graph Neural Networks for Graph-Level Supervised Learning** Ann Arbor, MI
  UMich Foreseer Group; Research Assistant Instructed by Prof. Qiaozhu Mei *October 2019 – December 2020*
  - Research Summary: Propose a novel fuzzy fingerprinting-based graph neural network (FF-GNN) to improve graph fingerprinting methods by parameterizing with a multi-channel fuzzy histogram, which has better trade-off in prediction accuracy, amortized time complexity, and GPU memory over other graph fingerprinting methods and MP-GNNs.
  - Developed smooth fuzzy histograms with PyTorch to convert node-level fingerprints to universal graph representations.
  - Visualized hidden layers and graph structures with TensorBoard and NetworkX to discover the performance bottlenecks.
  - Designed and ran all the experiments on synthetic and real-world datasets using lab's CUDA machines and AWS EC2, to make a fair comparison of accuracy and computational cost between FF-GNN and baseline models.
  - Wrote the first draft of Introduction, Approach and Experiment sections of the paper, and delivered an individual presentation to the research group summarizing the paper and research work.
- **Physical Causality in Grounded Word Acquisition** Ann Arbor, MI
  UMich Situated Language & Embodied Dialogue Group; Research Assitant Instructed by Prof. Joyce Y. Chai *May 2020 – Present*
  - Research Summary: Motivated by cognitive studies on children's language learning, discovered understanding physical causality of actions brings verb learning and noun learning in a bootstrap loop with 5.2% — 13.3% improvement.
  - Augmented the ACT dataset by adding human-labeled bounding boxes with annotations and cleansed the data.
  - Designed and implemented two noun grounding modules by adapting Faster R-CNN networks and attention networks, to determine bounding boxes with inputs of images, captions and noun embeddings from ResNet, BERT and word2vec.
  - Wrote the first draft of Related Work section of the paper and came up with multiple suggestions to other parts.

PROJECT EXPERIENCE

- **Iterative Data Processing with Spark** *December 2020*
  - Analyzed a Twitter social graph with Spark RDD and DataFrame on Zeppelin Notebook, and implemented the PageRank algorithm to find the most influential users.
  - Debugged the programs utilizing YARN & Spark UI and identified performance bottlenecks through DAG visualization.
  - Deployed the Spark applications to Azure Databricks and reduced the execution time by 12% compared to HDInsight.
- **Projects for EECS 484 Database Management Systems** *January 2020 – April 2020*
  - Designed the relational database schema for a Facebook-like service and developed a Java application for SQL executions with JDBC, to support features like user information queries, nearby event discoveries, and friend suggestions.
  - Migrated the Facebook dataset from Oracle SQL*Plus to MongoDB for higher schema flexibility in JSON format, and rewrote the queries in JavaScript.
- **Breast Cancer Dataset Analysis - STATS 415 Data Mining Group Project** *November 2019 – December 2019*
  - Based on "Breast Cancer Wisconsin Data Set" which recorded the characteristics of the cell nuclei, applied principal component analysis to explore the most informative combination of predictors.
  - Compared performance of logistic regression, random forest, SVM, LDA and QDA implemented in R, and found QDA with best subset selection is most effective to make a well-performed prediction for the diagnosis.
- **Projects & Labs for VE 280/EECS 281 Data Structures and Algorithms** *May 2019 – December 2019*
  - Utilized inheritance and basic dynamic polymorphism to implement sorted priority queue, binary heap priority queue and pairing heap priority queue developed from templated generic code.
  - Applied branch and bound algorithm and MST to solve the TSP problem for complete weighted graph.
  - Implemented a C++ version of 2048 which responded to player's keystrokes, and supported customized tiles like emojis.

COMPUTER SKILLS

- **Languages**: Python, C++, R, Scala, Java, SQL, MATLAB
- **Tools**: PyTorch, TensorFlow, scikit-learn, MySQL, SQL*Plus, MongoDB, Spark, AWS, Azure, Linux, Docker, Git