陈昊炜 (Eric Chen)

手机：(86)18917510907　邮箱：18917510907@189.cn　微信：ericchenhaowei

教育经历

复旦大学·信息科学与工程学院 | 主修专业：电子信息 | 第二专业：经济学 2019.09–至今

- 2019–2020年复旦大学优秀学生奖学金三等奖

复旦大学管理学院 | 跨专业学程项目：统计学&数据智能

- 主要课程：统计推断/回归分析/多元分析/抽样调查/时间序列分析/SQL数据库与企业数据管理/社交网络挖掘/计量经济学

海外交流经历

University of California, San Diego/加州大学圣地亚哥分校 | 学习领域：Computer & Data Science 2022.08–2023.01

- 主要课程：Discrete Mathematics/Deep Learning/Data Management/Algorithm Design & Analysis

实习经历

SAP中国研究院　数据建模与预测分析实习生 2023.02–至今

- 在Data Intelligence and Innovation团队，深入了解客户需求，从SAP HANA数据仓库中提取和处理大量数据进行**数据建模与分析。**
- 熟练掌握**ETL数据开发**流程，独立完成从数据抽取、转换到加载的全过程，并持续进行数据监测，确保数据质量和准确性。
- 利用SAP Datasphere等BI平台，定义与呈现深度业务指标，搭建3个不同板块的定制化看板(Dashboard)，为SAP实时监测企业需求提供极大支持。
- 应用**统计分析与机器学习**，在Azure ADB平台上进行KPI监测和商业**数据预测分析(Predictive Analysis)**，扩展ToB软件使用场景，提高企业运作效率。
- 持续参与企业AI Innovation相关案例研究，包括基于图算法的客户工单(Tickets)流程优化等，以技术助力团队。

美团·到店事业群　商业分析（平台机制与经济研究方向）实习生 2022.12–2023.02

- 负责到店餐饮团购单标准化的研究工作。通过**量化数据分析**，从平台视角出发推动美团平台内商家的团购单进行标准化，增加产品区分度，优化资源配置，降低用户的决策选择成本，从而改善用户体验。
- 定义供给侧评估指标，识别特色、优质团购，找出其共性，从而了解自身和竞对在不同属性团购单上的优劣势，为业务发展提供关键见解。
- 使用**HiveSQL**从美团数仓中提取数据，并且借助**机器学习**进行聚类，挖掘，对商家产品营销和用户消费行为进行洞察，寻找商家团单供给和用户团购需求偏好，为商家的团单优化设计与平台侧的推荐和搜索机制提供建设性指引。

贝塔斯曼欧唯特(中国)·客户关系管理(CX-LAB)　商业分析实习生 2022.05–2022.09

- 企业桌面研究，开展对屈臣氏，飞书，缤客等企业的桌面研究，寻求其客户关系管理间存在的痛点，最终以**脑图/分析报告**等形式呈现。
- 优化客户之声(Voice of Customer)自动分析系统。基于大量客户数据，建立精细化标签模型，以报告形式协助客户企业理解用户需求，提升客户体验。
- 使用定量分析手段，基于大数据对金融保险企业项目进行**客户画像分析。**通过深入洞察用户行为特征，为企业提供有针对性的客户细分策略。
- 围绕客户标签，建立深度学习模型进行**精准营销**，提升客户满意度，降低客户流失率与投诉率。

埃森哲（中国）　云服务项目组咨询顾问 2021.07–2021.08

- 在云服务技术咨询项目组，为上游云提供商与下游客户企业提供联结，挖掘企业内在需求，为企业客户设计云网络结构方案以**协助企业完成业务上云以及数字化转型**，并熟悉咨询公司的工作流程。
- 企业数据整理和分析，参与企业资源管理系统设计，同时为客户进行**企业风险评估和行业分析**。
- 项目支持工作，包括**Visio绘图，PPT制作和翻译**等。

研究与项目经历

图神经网络(Graph Neural Network)结构攻击与数据挖掘提取 — 复旦DISC实验室 2022.10–2023.05

- **在图数据挖掘领域内创新性地提出一种在图预训练模型上还原图结构数据集的方法，基于模型对生成的潜在子图结构进行合筛选与预测。该策略可有效应对多种预训练模型，并且弥补现有的数据提取方法的缺陷，有助于对于图隐私保护的进一步关注与研究。**
- 基于**Pytorch框架**，根据不同预训练模型(Contrastive learning/context prediction/masked component modeling)编写算法主要代码，并引入RL等多种生成结构筛选方法，提升模型表现。模型AUC相较于化学分析方法与常规机器学习基线测试提升30%以上。
- 设计**模型优化策略**，正则化参数以降低数据集分布对于模型的影响，增强模型的泛化性与鲁棒性。
- **在Linux服务器**上利用CUDA对模型开展实验与测试，对实验结果进行**数据分析与可视化**。并且编写论文。
- 论文已经投稿至国际顶级机器学习会议NeurIPS 2023。

基于Double Machine Learning对图网络进行因果推断(Causal Inference) — UC San Diego Database Lab 2022.09–至今

- **结合DML方法对图数据进行因果推断，并在有干扰的情况下估计网络的因果效应。此方法和情境旨在揭示社会网络中错综复杂的联系以及不同处理因素（如政策、药物）对现实结果（如经济增长、公共卫生）的影响，为政府和商业决策提供数据支持。**
- 深入研究图结构下有干扰的因果效应理论，协助课题组为实验和算法实现奠定理论基础。
- 基于不同图表示学习方法与图神经网络架构，优化数据生成过程与因果推断算法，大幅**提升模型学习率**，降低估计偏差。
- 设计对比实验，对模型进行性能微调，呈现试验结果。

图预训练模型数据增强(Data Augmentation)与隐私保护 — 复旦大学 2022.05–2023.01

- **提出一种注重隐私保护的图数据增强策略。在图预训练(Graph Pre-training)过程中根据图结构因果联系同时兼顾隐私保护(Privacy-preserving)与模型泛化性(generalization)。从而在模型共享的下游任务中更加有效地保护模型所有者的隐私数据信息。**
- 参与完善代码，并且完成实验，测试与结果分析，完成论文。论文已经投稿至NeurIPS 2023。

校园经历

复旦大学学生合唱团 2019.09–至今

- 2021年·成都 | 全国大学生艺术展演金奖
- 组织，策划并参与校内多场音乐会与演出活动，并受到观众的一致好评。

技能

C, Python, R, SQL, Spark, Microsoft Office, Tableau

数据库，数据分析，洞察与可视化，并且擅长在工作，科研与学习中积极利用AI技术大幅提升效率

Linux，算法，机器学习与Pytorch深度学习

流利英语能力（托福107分，大学英语六级624分），良好的沟通技能与合作能力

Haowei Chen (Eric)

Phone: (86)18917510907 E-mail: 18917510907@189.cn Wechat: ericchenhaowei

Education

Fudan University, Shanghai, China | Major: Electronic Engineering | Second Major: Economics **Sept 2019-Present**

- Third-Class Award for Scholarship of Excellent Student.

School of Management, Fudan University | Interdisciplinary: Statistics/Data-based Business Analysis

- Main Courses: Statistics Inference/Regression Analysis/Multivariate Analysis/Time Series/SQL/Network Mining/Econometrics

Overseas Experience

University of California, San Diego | Field of Study: Computer Science and Engineering **Aug 2022-Dec 2022**

- Main Courses: Discrete Mathematics/Deep Learning/Data Management/Algorithm Design & Analysis

Intern Experience

SAP Labs China **Data Modeling and Predictive Analysis Intern** **Feb 2023-Present**

- With the Data Intelligence and Innovation team, understand customer demands and process large volumes of data from the SAP HANA data warehouse for **data modeling and analysis**.
- Proficiently master the **ETL data development** process, independently completing the entire process from data extraction, transformation, to loading, and continuously monitoring the data to **ensure data quality and accuracy**.
- Utilize SAP Datasphere and other BI platforms to define and present in-depth business metrics, **build three customized dashboards for different sectors**, providing significant support to SAP's real-time monitoring of enterprise requirements.
- Apply **statistical analysis and machine learning** techniques on the Microsoft Azure ADB platform for KPI monitoring and business data **predictive analysis**, expanding ToB software use cases and enhancing enterprise operational efficiency.
- Continuously participating in AI Innovation case studies, including **graph algorithm-based** ticket component optimization, to utilize cutting-edge technology.

Meituan Corp. **Platform Mechanism and Economic Research Business Analytics Intern** **Dec 2022-Feb 2023**

- Responsible for conducting research on **standardization of group purchase orders** for in-store dining services. By utilizing **quantitative data analysis**, promoted standardization of group purchase orders within the Meituan platform, increasing product differentiation, optimizing resource allocation, and reducing user decision-making costs, ultimately improving user experience.
- **Defined supply-side evaluation metrics**, identified distinctive and high-quality group purchases, and discovered their common features to gain insights of both ourselves and competitors in various group purchase order attributes, providing critical insights for business development.
- **Employed HiveSQL** to extract data from the Meituan data warehouse and **applied machine learning algorithms for clustering and mining**, enabling insights into merchant and customer behavior. Identified their preferences, and offered constructive guidance for group purchase optimization and platform-side recommendation and search mechanisms."

Bertelsmann Arvato Corp. **CRM Dept. (CX-Lab) Business Analytics Intern** **May 2022-Sept 2022**

- **Conducted desktop research** on companies such as Watsons, Lark, and booking.com, identifying pain points in their customer relationship management and presenting findings in the form of mind maps and analytical reports.
- Optimize the analysis system *Voice of Customer*. Based on a large volume of customer data, established granular labeling model to assist customers in understanding user needs and improving experience.
- Quantitative analysis to perform customer profiling analysis for financial and insurance enterprise projects. By gaining insights into user behavior characteristics, provided targeted customer segmentation strategies for businesses.
- Built deep learning models around customer tags for precision marketing, enhancing customer satisfaction, and reducing customer loss and complaint rates.

Accenture Inc. **Project Team Assistant Consultant** **July 2021-Aug 2021**

- As a cloud service technical consulting project team for enterprise customer, consulting services are provided to explore their inherent needs, and designed cloud network structure for **customer's digital transformation.**
- Enterprise data organization and analysis, participation in enterprise resource management system design, as well as enterprise risk assessment and industry analysis.
- Project support work, including Visio drawing, PPT production and translation, etc.

Research Experience

Graph Data Extraction from Graph Pre-trained Model — Fudan DISC Lab **Oct 2022-May 2023**

- **Propose an innovative method in the field of graph data mining on graph pre-trained models to restore graph structure datasets, based on the model's screening and prediction of generated potential subgraph structures. This strategy can effectively address various pre-trained models and compensate for the shortcomings of existing data extraction methods, contributing to graph privacy protection.**
- Implement main algorithm codes for different pre-trained models (Contrastive learning/context prediction/masked component modeling) based on the Pytorch framework, and introduce various generation screening methods including Reinforcement Learning (RL) Generation to improve model performance. The model's AUC is improved by over 30% compared to chemical analysis methods and conventional machine learning baseline tests.
- Design optimization strategies and regularizations to reduce the impact of dataset distribution on the model, enhancing the model's generalization and robustness.
- Conducted experiments and tests on Linux servers using CUDA, performed data analysis on experimental results, and visualization. Co-authored the research paper.
- The paper has been submitted to the top international machine learning conference NeurIPS 2023.

GNN based Double Machine Learning Estimator of Network Causal Effects — UCSD Database Lab **Sept 2022-Present**

- **Combined Double Machine Learning methods to perform causal inference on graph data and estimate the causal effects of networks under interference. This method aims to reveal the intricate relationships in social networks and the impact of different treatment factors (e.g., policies, drugs) on real-world outcomes (e.g., economic growth, public health), providing data support for government and business decision-making.**
- Conduct in-depth research on the theory of causal effects with interference in graph structures, and assist the research group in laying the theoretical foundation for experiments and algorithm implementation.
- Optimize data generation processes and causal inference algorithms based on different graph representation learning methods and graph neural network architectures, significantly improving model learning rates and reducing estimation biases.
- Design comparative experiments, fine-tuned model performance, and presented experimental results.

Release Graph Pre-trained Model via Privacy-Preserving Data Augmentation — Fudan University **May 2022-Jan 2023**

- **Propose a privacy-preserving graph data augmentation strategy. During the graph pre-training process, privacy preservation and model generalization were considered according to the causal relationships of graph structures. This protects the data privacy of model owners in downstream tasks of model sharing.**
- Participated in refining the code, completing experiments, tests, and result analysis, and finalizing the paper. The paper has been submitted to NeurIPS 2023.

Campus Experience

Fudan University Student Choir

- 2021 · Chengdu | National College Student Art Exhibition **Gold Award** **Sept 2019-Present**
- Organized, planned, and participated in multiple concerts and performances on campus, receiving unanimous praise from the audience.

Skills and Interest

C, Python, R, SQL, Spark, Microsoft Office, Tableau.
Database, data analytics, business insights and visualization. Proficient in utilizing AI to improve efficiency in work, research, and learning.
Linux, Algorithm, background and experience of machine learning and deep learning with Pytorch
Fluent English(TOEFL 107), with good communication skills and collaboration.