

邓紫臻

✉ dengzizhen2021@ia.ac.cn · 📞 (+86) 18803835618

🎓 教育背景

中国科学院自动化研究所，社会计算专业，硕士在读	2021 – 2024
研究方向：因果推理，因果学习，模型对齐，信息级联，社会弹性	
华北电力大学，计算机科学与技术专业，工学学士学位	2017 – 2021
GPA 91.46 (1/119)，国家奖学金，校一等奖学金，校三好学生标兵，优秀毕业生	

🔬 科研/项目经历

1. 大语言模型的因果对齐 2023.05 – 2023.07

现有的可解释性工具在数亿参数的大语言模型上通常无法适配，因为它们往往专注于为特定任务微调的小型模型。在本研究中，我们提出了一种基于因果抽象理论和混合专家模型的新方法，以找到在模型中对应给定因果变量的神经表示。亮点如下：

- 将因果推理中的干预操作引入大语言模型，利用设计的不同输入实现对中间变量的软干预，通过实际输出与模型输出的误差来训练使得大语言模型中的神经表示与因果模型中的因果变量对齐
- 就我们所知，此工作 第一个 将混合专家用于大模型的因果对齐，探索混合专家模型的解耦功能，非线性的分离神经表示中的高维语义因果变量，与人类认知对齐，使得模型提供更好的可解释性。

2. 基于因果发现的社会弹性研究 2022.02 – 2023.05

社会弹性揭示了社会面临风险和压力时的演化趋势，以往的社会弹性度量方法存在较大的主观性和不确定性。本研究借助因果发现生成的社会认知网络的弹性来表征社会弹性，将社会弹性当做一个整体而不是割裂的指标。亮点如下：

- 我们利用 LIWC 提取社交媒体文本中的认知维度，使用因果发现方法寻找认知维度之间的因果关系，构成社会认知网络。使用 GBB 算法将网络弹性的多维结构简化为一维结构，从而可以进行解析描述。
- 就我们所知，此工作 第一个 将认知网络引入社会弹性的研究来捕捉人们认知的变化，很大程度上避免了主观性，具有更高的可解释性。

3. 基于因果推理的级联行为建模 2022.09 – 2023.02

目前的信息级联以预测为主，忽略了对个体级联行为的理解。本研究利用因果推理，计算个体决策时分别受内在动机和外在环境的影响程度，理解个体行为背后的原因，为政策制定提供依据。亮点如下：

- 利用结构因果模型中的对撞结构，参数化用户内在动机和外在环境，在已知级联行为的情况下，利用误差训练补全缺失表示。
- 就我们所知，此工作是 第一个 在信息缺失的情况下实现对个体的级联行为进行理解的尝试。

4. 深度因果学习：表征，发现和推理 2021.11 – 2022.09

本研究指出现在因果学习领域存在的问题与局限，引出与深度学习结合的必然性。然后依次介绍面向因果变量的表征学习、深度因果发现、深度因果推理这三个核心方向最新的方法。亮点如下：

- 就我们所知，此工作 第一个 全面综述了因果表征、因果发现、因果推理与深度学习的结合。
- 包含经典的和最新的深度因果发现和深度因果推理方法，如 NOTEARS, DAG-GNN, Grad-DAG; CFR, DragonNet, CEVAE, GANITE, DeepIV 等
- 发表论文：Deep Causal Learning: Representation, Discovery and Inference.
- 一作，ACM Computing Survey (under review), ACM 综述类旗舰期刊, arXiv:2211.03374

5. 因果推理在社交媒体中的应用

2021.06 – 2021.11

本研究基于大规模新闻媒体数据，利用交叉收敛映射方法构建国家之间的影响力因果网络，叠加交互出入度生成的影响力网络形成可信影响力网络，结果具有更高的可靠性与可解释性。亮点如下：

- 就我们所知，此工作 **第一个** 将因果推理用于世界范围新闻媒体，度量国家之间因果影响力
- 通过构建的可信影响力网络，我们能够得到哪些国家通过自身影响力对其他国家产生影响
- 发表论文：Credible Influence Analysis in Mass Media Using Causal Inference.
- 一作，IEEE ISI 2021, EI, 安全信息学顶级会议

🏆 比赛经历

PCIC 第二届华为因果推理挑战赛

2022.09 – 2022.12

- 使用 TCN, CNN+Transformer 等方法，使不同的模型块捕捉不同时间的异常特征，最大程度提高模型的泛化性
- 方案准确率排名第 5，获得铜奖，受邀参加 2022 PCIC Competition Workshop

渔船作业方式识别—DataFountain 训练赛

2021.11 – 2021.12

- 构造特征，使用随机森林、XGBoost、CNN、LSTM 等多种分类器实现作业方式识别
- 准确率达到 0.944，排名 21/396

天猫复购预测—阿里云天池挑战赛

2021.10 – 2021.11

- 对数据进行清洗、可视化，从商家、用户等多个角度进行特征工程构造新的特征
- 使用 XGBoost、LightGBM 和 CatGBM 三个模型进行训练预测，取得前 10% 的效果

其他比赛奖项

2017-2021

- 国家级大学生创新创业优秀项目负责人
- 国际水中机器人大赛二等奖
- 中国大学生数学建模竞赛省二等奖
- 美国大学生数学建模竞赛 H 奖
- 华北电力大学“互联网+”大赛金奖
- 华北电力大学计算机应用大赛二等奖

⚙️ 技能

- 编程语言：Python > C++
- 英语水平：CET-4 (530), CET-6 (507)
- 深度学习：掌握 TensorFlow, PyTorch 等深度学习框架
- 因果学习：掌握 YLearn, DoWhy, CausalML, EconML, TrustworthyAI 等因果学习框架

💎 自我评价

- 基础扎实：研究生期间学习了机器学习、深度学习、数据挖掘领域的经典模型算法。深入研究了因果学习领域，包括因果表征，因果推理，因果发现，特别是因果学习与深度学习的结合。对于因果科学领域有全面深入的了解，清楚发展脉络与存在问题，对未来方向有一定的感知。
- 学科交叉：作为社会计算专业学生，拥有多学科视野，同时涉猎包括信息级联、社会弹性、情感分析、复杂网络、实证分析等多个研究领域。能从多个角度看问题，擅长交叉创新。
- 志向坚定：首先，深入研究人工智能理论，做出技术上的突破；之后，将先进的人工智能技术应用于社会场景中，为人们生活带来切实的改善。