Copula 熵的多学科应用

马健博士
majian03@gmail.com

2023 年 10 月 5 日

Contents

Copula 熵

Jian Ma and Zengqi Sun. "Mutual Information Is Copula Entropy". In: *Tsinghua Science & Technology* 16.1 (2011). See also arXiv preprint arXiv:0808.0845 (2008), pp. 51–54

$$H_c(\mathbf{x}) = -\int_{\mathbf{u}} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u} \tag{1}$$

综述论文

马健. “Copula 熵：理论和应用”. In: *ChinaXiv:202105.00070* (2021)

摘要

统计独立性是统计学和机器学习领域的基础性概念，如何表示和度量统计独立性是该领域的基本问题。Copula 理论提供了统计相关关系表示的理论工具，而 Copula 熵理论则给出了度量统计独立性的概念工具。本文综述了 Copula 熵的理论和应用，概述了其基本概念定义、定理和性质，以及估计方法。介绍了 Copula 熵研究的最新进展，包括其在统计学的九个基本问题（结构学习、关联发现、变量选择、因果发现、系统辨识、时延估计、域自适应、正态性检验和双样本检验等）上的理论应用。讨论了前四个理论应用之间的关系，以及其对应的深层次的相关性和因果性概念之间的联系，并将 Copula 熵的（条件）独立性度量框架与基于核函数和距离相关的同类框架进行了理论对比，又通过仿真和实际数据实验评估验证了 Copula 熵的实际优越性。简述了 Copula 熵在理论物理学、理论化学、化学信息学、材料学、水文学、气候学、气象学、环境学、生态学、动物形态学、农学、认知神经学、运动神经学、计算神经学、心理学、系统生物学、生物信息学、临床诊断学、老年医学、精神病学、公共卫生学、经济学、管理学、社会学、教育学、计算语言学、新闻传播学、法学、政治学、军事学、情报学，以及能源工程、食品工程、土木建筑、交通运输、制造工程、可靠性工程、化学工程、航空航天、车辆工程、电子工程、通信工程、高性能计算、测绘遥感和金融工程等领域的实际应用。

理论物理学

- 相关粒子系统
  - 平衡态相关粒子系统中熵的推导和计算[1]

[1] Jian Ma. "On Thermodynamic Interpretation of Copula Entropy". In: *arXiv preprint arXiv:2111.14042* (2021). arXiv: 2111.14042 [cs.IT].

理论化学

- 变构效应研究
  - 变构效应配位点和激活点热力学耦合模型[2]
    - 丙氨酸二肽的 C 端和 N 端

[2] Michel A. Cuendet, Harel Weinstein, and Michael V. LeVine. "The Allostery Landscape: Quantifying Thermodynamic Couplings in Biomolecular Systems". In: *Journal of Chemical Theory and Computation* 12.12 (Dec. 2016), pp. 5758–5767. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.6b00841. URL: https://doi.org/10.1021/acs.jctc.6b00841.

化学信息学

- 分子设计
  - 设计具有特定属性的分子结构[3]
    - 有机分子属性 QM9 数据库

[3] Mario Wieser. "Learning Invariant Representations for Deep Latent Variable Models". PhD thesis. University of Basel, 2020.

材料学

- 耐热型含能材料分子设计
  - 含能材料热分解温度预测[4]
    - 含能材料化合物分子结构和量子化学性质数据集

[4] 田杰. “基于机器学习的耐热型含能材料设计方法研究”. 硕士学位论文. 西南科技大学, 2023.

水文学 I

- 洪水预报
  - 金沙江流域洪水预报[5]
- 河流相关性
  - 长江上游河段（金沙江、岷江、沱江、嘉陵江）相关性[6]
  - 长江上游河段复合洪水事件分析[7]
- 水沙关系分析
  - 黄河西柳沟河流域径流量和输沙量数据分析[8]

[5] Lu Chen, Vijay P. Singh, and Shenglian Guo. "Measure of Correlation between River Flows Using the Copula-Entropy Method". In: *Journal of Hydrologic Engineering* 18.12 (2013), pp. 1591–1606, Lu Chen et al. "Copula entropy coupled with artificial neural network for rainfall–runoff simulation". In: *Stochastic Environmental Research and Risk Assessment* 28.7 (2014), pp. 1755–1767.

[6] Lu Chen and Shenglian Guo. *Copulas and its application in hydrology and water resources*. Springer, 2019.

[7] Xu Wang and Yong-Ming Shen. "A Framework of Dependence Modeling and Evaluation System for Compound Flood Events". In: *Water Resources Research* 59.8 (2023), e2023WR034718. DOI: https://doi.org/10.1029/2023WR034718.

[8] Longxia Qian et al. "A New Estimation Method for Copula Parameters for Multivariate Hydrological Frequency Analysis With Small Sample Sizes". In: *Water Resources Management* 36.4 (Mar. 2022), pp. 1141–1157. ISSN: 1573-1650. URL: https://doi.org/10.1007/s11269-021-03016-w.

水文学 II

- 干旱研究
  - 黄河流域 (河南和甘肃) 干旱分析和预测[9]
  - 印度达布蒂（Tapti）河流域干旱指数设计[10]
  - 伊朗城市（扎黑丹、恩泽利和马什哈德）干旱数据分析[11]
- 水文事件风险建模
  - 黄河流域干旱事件识别[12]
- 中长期径流预报
  - 长江上游水文站月径流预测[13]

---

[9] 温云亮 et al. “基于 Copula 熵理论的干旱驱动因子选择”. In: 华北水利水电大学学报 (自然科学版) 40.4 (2019), pp. 51–56, C.Y. Huang and Y.P. Zhang. “Prediction based on Copula Entropy and General Regression Neural Network”. In: *Applied Ecology and Environmental Research* 17.6 (2019), pp. 14415–14424, 黄春艳. “黄河流域的干旱驱动及评估预测研究”. 博士学位论文. 西安理工大学, 2021.

[10] P. Kanthavel, C.K. Saxena, and R.K. Singh. “Integrated Drought Index based on Vine Copula Modelling”. In: *International Journal of Climatology* (2022). DOI: https://doi.org/10.1002/joc.7840. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.7840. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.7840.

[11] M. Mohammadi, M. Emadi, and M. Amini. “Bivariate Dependency Analysis using Jeffrey and Hellinger Divergence Measures based on Copula Density Estimation by Improved Probit Transformation”. In: *Journal of Statistical Sciences* 15.1 (2021), pp. 233–254. DOI: 10.29252/jss.15.1.233.

[12] Lingling Ni et al. “Vine copula selection using mutual information for hydrological dependence modeling”. In: *Environmental Research* 186 (2020), p. 109604.

[13] Xiao Li et al. “Predicting Monthly Runoff of the Upper Yangtze River Based on Multiple Machine Learning Models”. In: *Sustainability* 14.18 (2022), p. 11149. ISSN: 2071-1050. DOI: 10.3390/su141811149. URL: https://www.mdpi.com/2071-1050/14/18/11149.

水文学 III

- 中长期径流预报
  - 南水北调工程丹江口水库入库径流预报[14]
  - 洪泽湖和骆马湖长期径流预报[15]
- 流域分区[16]
- 多站点径流生成
  - 巴西雅瓜拉比-大都市水库系统[17]

[14] 黄朝君 et al. “基于 Copula 熵-随机森林的中长期径流预报研究”. In: 人民长江 52.11 (2021), pp. 81–85.

[15] Ran Mo et al. “Long-term probabilistic streamflow forecast model with “inputs–structure–parameters” hierarchical optimization framework”. In: *Journal of Hydrology* (2023), p. 129736. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2023.129736. URL: https://www.sciencedirect.com/science/article/pii/S0022169423006789.

[16] 刘磊 et al. “基于非线性相关性和复杂网络的径流相似性分区”. In: 水科学进展 33.3 (2022), pp. 442–451.

[17] Victor Costa Porto et al. “A GLM copula approach for multisite annual streamflow generation”. In: *Journal of Hydrology* 598 (2021), p. 126226.

水文学 IV

- 水文观测网络选址和优化
  - 上海雨量观测网[18]
  - 伊洛河流域水文观测网[19]
  - 北京市区水文观测网; 汾河流域观测网; 太湖盆地流域雨量观测网[20]
  - 淮河流域雨量观测网[21]
  - 美国查克托哈奇（Choctawhatchee）河流域水文观测网[22]

[18] Pengcheng Xu et al. "A two-phase copula entropy-based multiobjective optimization approach to hydrometeorological gauge network design". In: *Journal of Hydrology* 555 (2017), pp. 228–241.

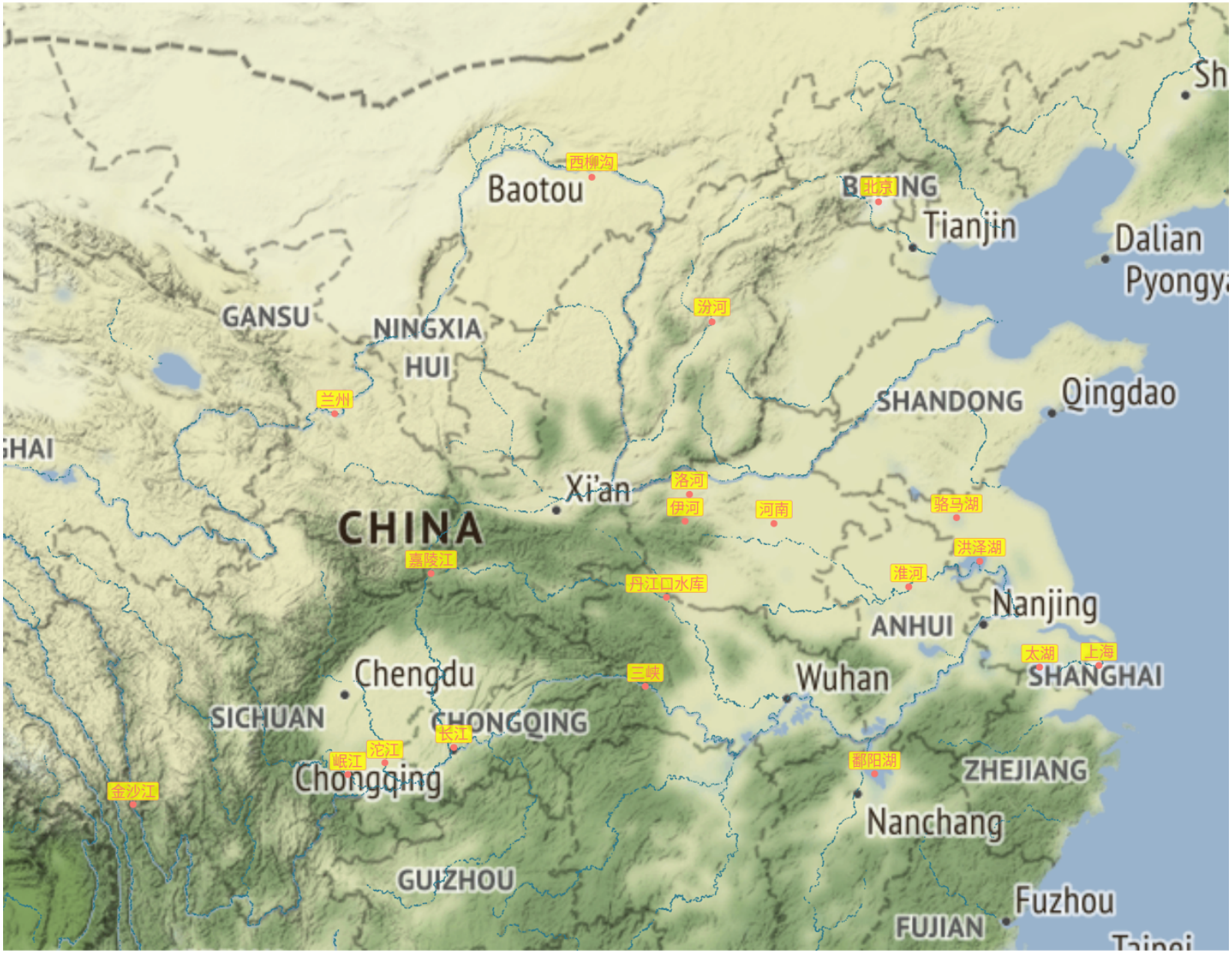
[19] 王栋 et al. "一种基于 Copula 熵的水文站网优化模型的优化方法". Pat. CN106897530B. CN106897530B. 2019.

[20] Heshu Li et al. "Developing a dual entropy-transinformation criterion for hydrometric network optimization based on information theory and copulas". In: *Environmental Research* 180 (2020), p. 108813.

[21] 徐鹏程 et al. "基于 C-Vine Copula 熵多目标优化模型的水文气象站网优化研究". In: 中国农村水利水电 2 (2022), pp. 16–21. DOI: 10.12396/znsd.220702.

[22] 杨惜岁. "多目标准则下流域水文站网的优化与评价". 硕士学位论文. 武汉理工大学, 2019.

水文学-国内应用地图


Baotou
西柳沟
Tianjin
北京
Dalian
GANSU
NINGXIA
HUI
兰州
汾河
SHANDONG
Qingdao
Xi'an
洛河
伊河
河南
CHINA
骆马湖
洪泽湖
嘉陵江
丹江口水库
淮河
Nanjing
ANHUI
Chengdu
三峡
Wuhan
太湖
上海
SHANGHAI
SICHUAN
CHONGQING
长江
沱江
岷江
Chongqing
金沙江
鄱阳湖
ZHEJIANG
Nanchang
GUIZHOU
Fuzhou
FUJIAN

气候学

- 气候变化对水文气象变量相关性的影响
  - 美国德州达拉斯地区降水和气温相关性关系分析[23]
    - 达拉斯市沃斯堡（Fort Worth）地区降水和气温观测数据
- 气候评估
  - 欧洲城市气候分类[24]
    - 欧洲 25 座城市气温和降水数据

[23] Zengchao Hao and Vijay P. Singh. "Integrating Entropy and Copula Theories for Hydrologic Modeling and Analysis". In: *Entropy* 17.4 (2015), pp. 2253–2280. ISSN: 1099-4300. DOI: 10.3390/e17042253. URL: https://www.mdpi.com/1099-4300/17/4/2253.

[24] Francesca Condino. "La divergenza di Jensen-Shannon nell'algoritmo di clustering dinamico per dati descritti da distribuzioni multivariate". PhD thesis. Università degli Studi di Napoli Federico II, 2009.

气象学

- 大气污染气象成因分析
  - 北京地区气象因素对 PM2.5 浓度的因果关系分析[25]
    - 北京地区 PM2.5 和气象观测数据
  - PM2.5 浓度预测[26]
    - 北京地区 PM2.5 和气象观测数据
  - 上海和广州大气污染预测预警[27]
    - 上海和广州 PM2.5 和气象观测数据
- 气象灾害预测
  - 广西地区台风灾情预测类[28]
    - 广西地区台风灾害数据

---

[25] Jian Ma. "Estimating Transfer Entropy via Copula Entropy". In: *arXiv preprint arXiv:1910.04375* (2019).

[26] Xiaoxuan Wu et al. "Research on PM2.5 Concentration Prediction Based on the CE-AGA-LSTM Model". In: *Applied Sciences* 12.14 (2022), p. 7009. ISSN: 2076-3417. DOI: 10.3390/app12147009. URL: https://www.mdpi.com/2076-3417/12/14/7009, Jieyin Chen. "Short-Term Prediction of PM2.5 Concentration based on Self-Attention Mechanism Improved Temporal Convolution Network". In: *2023 International Seminar on Computer Science and Engineering Technology (SCSET)*. 2023, pp. 528–534. DOI: 10.1109/SCSET58950.2023.00121.

[27] Jujie Wang et al. "A novel air quality prediction and early warning system based on combined model of optimal feature extraction and intelligent optimization". In: *Chaos, Solitons & Fractals* 158 (2022), p. 112098. ISSN: 0960-0779. DOI: https://doi.org/10.1016/j.chaos.2022.112098. URL: https://www.sciencedirect.com/science/article/pii/S0960077922003083.

[28] 陈燕璇, 刘合香, and 倪增华. "基于 Copula 熵因子选取的 PSO-ELM 台风灾情预测模型". In: 气象研究与应用 40.2 (2019), pp. 7–11.

环境学

- 大气污染传播路径分析
  - 兰州市大气污染传播路径预测[29]
    - 兰州市环境气象检测网络 2017 年 PM2.5 观测数据
- 火电厂排放污染物管控
  - 火电厂氮氧化物排放浓度预测[30]
    - 宁夏某燃煤电厂运行数据

[29] 吴京鹏. "基于图嵌入表示的节点无特征网络链路预测研究". 硕士学位论文. 西北师范大学, 2022.

[30] 金秀章, 乔鹏, and 史德金. "基于 VMD-Bayes-Lasso 算法带误差补偿的火电厂 NOx 浓度软测量". In: 华北电力大学学报 (自然科学版) (2023).

生态学

- 动物运动轨迹分析
  - Cylcop 算法包[31]

[31] Florian H. Hodel and John R. Fieberg. "Cylcop: An R Package for Circular-Linear Copulae with Angular Symmetry". In: *bioRxiv* (2021), p. 2021.07.14.452253, Florian Hodel. *cylcop: Circular-Linear Copulas with Angular Symmetry for Movement Data*. CRAN. R package version 0.2.0, URL: https://cran.r-project.org/package=cylcop. 2022. URL: https://cran.r-project.org/package=cylcop.

动物学

- 动物形态学
  - 鱼类形态相似度研究[32]
    - GatorBait 海洋鱼类外形数据库
  - 鲍鱼生长过程的形态学研究[33]
    - UCI 鲍鱼数据集

[32] Francisco Escolano et al. "The mutual information between graphs". In: *Pattern Recognition Letters* 87 (2017), pp. 12–19. DOI: https://doi.org/10.1016/j.patrec.2016.07.012. URL: https://www.sciencedirect.com/science/article/pii/S016786551630174X.

[33] Soumik Purkayastha and Peter X.K. Song. "Asymmetric predictability in causal discovery: an information theoretic approach". In: *arXiv preprint arXiv:2210.14455* (2022).

农学

- 作物产量预测
  - 气候变化对我国南方两季稻产量的影响及对策[34]
    - 南方（江南和华南）54 个地点未来气候变化数据和作物数据

[34] Ziya Zhang et al. "Impact of climate change and planting date shifts on growth and yields of double cropping rice in southeastern China in future". In: *Agricultural Systems* 205 (2023), p. 103581. ISSN: 0308-521X. DOI: https://doi.org/10.1016/j.agsy.2022.103581. URL: https://www.sciencedirect.com/science/article/pii/S0308521X22002177.

神经科学 I

- 认知神经学
  - 分析大脑认知活动的多模态数据[35]
    - 人脸检测任务 EEG 数据
    - 听觉语音刺激任务和认知行为映射任务 MEG 数据
    - 奖惩学习任务前脑岛 (anterior Insula)SEEG 数据
  - 语音信息的编码和解析[36]
    - 故事讲述语音及相应的 EEG 数据
  - 因果关系脑连接网络分析[37]
    - 注意缺陷多动障碍患者 EEG 数据
    - 葡萄牙老年人静息态 fMRI 数据

---

[35] Stephanie J. Kayser et al. "Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha". In: *The Journal of Neuroscience* 35.44 (2015), pp. 14691–14701, Robin A. A. Ince et al. "The Deceptively Simple N170 Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres". In: *Cerebral Cortex* 26.11 (2016), pp. 4123–4135, Robin A.A. Ince et al. "A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula". In: *Human Brain Mapping* 38.3 (2017), pp. 1541–1573, Etienne Combrisson et al. "Group-level inference of information-based measures for the analyses of cognitive brain networks from neurophysiological data". In: *NeuroImage* (2022), p. 119347. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2022.119347. URL: https://www.sciencedirect.com/science/article/pii/S1053811922004669.

[36] Pieter De Clercq et al. "Beyond linear neural envelope tracking: a mutual information approach". In: *Journal of Neural Engineering* 20.2 (2023), p. 026007. DOI: 10.1088/1741-2552/acbe1d. URL: https://dx.doi.org/10.1088/1741-2552/acbe1d.

[37] Paolo Victor Redondo, Raphaël Huser, and Hernando Ombao. "Measuring Information Transfer Between Nodes in a Brain Network through Spectral Transfer Entropy". In: *arXiv preprint arXiv:2303.06384* (2023), 汪方毅 et al. "基于静息态 fMRI 区分健康老年人认知水平的 MVPA 方法研究". In: 磁共振成像 14.6 (2023), pp. 18–25.

神经科学 II

- 运动神经学
  - 分析运动的肌肉组合协同策略[38]
    - 伸手运动时肌肉 sEMG 数据
    - 自主运动肌肉疲劳状态 sEMG 数据
- 计算神经学
  - 神经元可塑性建模[39]
  - 神经网络信息传输关系分析[40]

[38] 吴亚婷 et al. "多尺度肌间耦合网络分析". In: 生物医学工程学杂志 38.4 (2021), pp. 742–752, Yating Wu et al. "R-Vine Copula Mutual Information for Intermuscular Coupling Analysis". In: *Proceedings of the 11th International Conference on Computer Engineering and Networks*. 2022, pp. 526–534, David Ó' Reilly and Ioannis Delis. "A network information theoretic framework to characterise muscle synergies in space and time". In: *Journal of Neural Engineering* 19.1 (2022), p. 016031. DOI: 10.1088/1741-2552/ac5150. URL: https://doi.org/10.1088/1741-2552/ac5150, Shaojun Zhu et al. "Intermuscular coupling network analysis of upper limbs based on R-vine copula transfer entropy". In: *Mathematical Biosciences and Engineering* 19.9 (2022), pp. 9437–9456, 金国美 et al. "基于小波包-Copula 互信息的肌间耦合特性". In: 传感技术学报 35.10 (2022), pp. 1348–1353.

[39] Johannes Leugering and Gordon Pipa. "A Unifying Framework of Synaptic and Intrinsic Plasticity in Neural Populations.". In: *Neural Computation* 30.4 (2018), pp. 945–986.

[40] Ari Pakman et al. "Estimating the Unique Information of Continuous Variables in Recurrent Networks". In: *Advances in Neural Information Processing Systems* (2021).

心理学

- 生物心理学
  - 情绪刺激下心跳诱发脑电位的时间交互现象[41]
    - 用于情绪分析的生理信号 DEAP 数据集

[41] Liesa Ravijts. "Revealing temporal interactions around the heartbeat-evoked potential modulated by emotional perception". MA thesis. Ghent Univeristy, 2019.

生物学

- 系统生物学
  - 生物信号调控和传导[42]
    - 癌症分子机制数据
  - 生物现象动态网络结构和功能[43]
    - 酵母细胞周期数据
- 生物信息学
  - 分析基因数据，研究生命和疾病机理[44]
    - 肝炎病毒感染治疗基因表达谱数据
  - 筛选与癌症有关的变异基因[45]
    - cBioPortal 癌症基因组数据
    - 亚利桑那州立大学癌症基因组数据

---

[42] Agata Charzyńska and Anna Gambin. "Improvement of the k-NN Entropy Estimator with Applications in Systems Biology". In: *Entropy* 18.1 (2015), p. 13.

[43] Farzaneh Farhangmehr et al. "An information-theoretic algorithm to data-driven genetic pathway interaction network reconstruction of dynamic systems". In: *2013 IEEE International Conference on Bioinformatics and Biomedicine*. 2013, pp. 214–217.

[44] Mario Wieser et al. "Inverse Learning of Symmetries". In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 18004–18015.

[45] Qiang Wu and Dongxi Li. "CRIA: An Interactive Gene Selection Algorithm for Cancers Prediction Based on Copy Number Variations". In: *Frontiers in Plant Science* 13 (2022), p. 839044. ISSN: 1664-462X. DOI: 10.3389/fpls.2022.839044. URL: https://www.frontiersin.org/article/10.3389/fpls.2022.839044.

医学 I

- 临床医学
  - 心脏病诊断[46]
    - UCI 心脏病数据
  - 糖尿病病情管理[47]
    - 美国 Health Facts 糖尿病救治网络数据
  - 癌症预后[48]
    - UCI 肺癌数据
  - 乳腺癌预后[49]
    - SEER 数据库乳腺癌临床数据
  - 白内障术后角膜水肿风险预测[50]
    - 临床白内障超声乳化手术患者数据
  - 主动脉瓣置换手术射血分数分析[51]
    - 临床主动脉瓣置换手术前后射血分数数据

[46] Jian Ma. "Variable Selection with Copula Entropy". In: *Chinese Journal of Applied Probability and Statistics* 37.4 (2021). See also arXiv preprint arXiv:1910.12389 (2019), pp. 405–420.

[47] Radko Mesiar and Ayyub Sheikhi. "Nonlinear Random Forest Classification, a Copula-Based Approach". In: *Applied Sciences* 11.15 (2021), p. 15. ISSN: 2076-3417. DOI: 10.3390/app11157140. URL: https://www.mdpi.com/2076-3417/11/15/7140.

[48] Jian Ma. "Copula Entropy based Variable Selection for Survival Analysis". In: *arXiv preprint arXiv:2209.01561* (2022).

[49] 付金露. "基于特征选择的乳腺癌患者预后模型研究". 硕士学位论文. 江西财经大学, 2023.

[50] Yu Luo et al. "Research on Establishing Corneal Edema after Phacoemulsification Prediction Model Based on Variable Selection with Copula Entropy". In: *Journal of Clinical Medicine* 12.4 (2023), p. 1290. ISSN: 2077-0383. DOI: 10.3390/jcm12041290.

[51] S.M. Sunoj and N. Unnikrishnan Nair. "Survival Copula Entropy and Dependence in Bivariate Distributions". In: *REVSTAT-Statistical Journal* (2023). URL: https://revstat.ine.pt/index.php/REVSTAT/article/view/560.

医学 II

- 认知医学
  - 认知能力评估 / 痴呆症筛查[52]
    - 北京和天津痴呆症老年人数据
- 运动医学
  - 运动能力评估 / 跌倒风险预测[53]
    - 天津和成都跌倒人群老年人数据
  - 重复经颅磁刺激对帕金森病改善神经机制分析[54]
    - 帕金森患者经颅磁刺激前后 EEG 数据
- 精神病学
  - 抑郁症患者识别[55]
    - 江苏常州抑郁症青少年患者 EEG 数据

---

[52] Jian Ma. "Predicting MMSE Score from Finger-Tapping Measurement". In: *Proceedings of 2021 Chinese Intelligent Automation Conference*. See also bioRxiv 817338 (2019). 2022, pp. 294–304. ISBN: 978-981-16-6372-7.

[53] Jian Ma. "Predicting TUG score from gait characteristics based on video analysis and machine learning". In: *Proceedings of 2023 Chinese Intelligent Automation Conference*. See also bioXiv 963686 (2020). 2023, pp. 1–12, Jian Ma. "Associations between finger tapping, gait and fall risk with application to fall risk assessment". In: *arXiv preprint arXiv:2006.16648* (2020).

[54] 李润泽 et al. "重复经颅磁刺激改善帕金森病运动症状的脑功能网络分析". In: 生物化学与生物物理进展 50.1 (2023), pp. 126–134.

[55] 张婷婷 et al. "基于 Couple 熵的抑郁症相干性反馈指标提取". In: 电子测量技术 45.9 (2022), pp. 160–167.

公共卫生学

- 新冠肺炎流行病 (COVID19)
  - 发热症状疑似病人筛查诊断[56]
    - 新冠临床数据
- 高血压
  - 高血压关联基因研究[57]
    - ELEMENT 数据集

[56] Radko Mesiar and Ayyub Sheikhi. "Nonlinear Random Forest Classification, a Copula-Based Approach". In: *Applied Sciences* 11.15 (2021), p. 15. ISSN: 2076-3417. DOI: 10.3390/app11157140. URL: https://www.mdpi.com/2076-3417/11/15/7140.

[57] Soumik Purkayastha and Peter X.K. Song. "Asymmetric predictability in causal discovery: an information theoretic approach". In: *arXiv preprint arXiv:2210.14455* (2022).

社会科学 I

- 经济学
  - 扶贫政策效果评估，用于政策目标人口鉴别[58]
    - 2018 年政府贫困家庭状况普查数据（四川省）
  - 议价机制中的互惠行为和时间效应[59]
    - eBay 的 Best Offer 平台数据
  - 产业链内部相关性分析[60]
    - 国内畜禽养殖产业链主要上市企业股票价格数据
  - 投资者情绪分析
    - 中国新能源汽车上市公司的百度搜索数据

[58] Qingsong Shan and Qianning Liu. "Binary Trees for Dependence Structure". In: *IEEE Access* 8 (2020), pp. 150989–150998. DOI: 10.1109/ACCESS.2020.3017529, 罗良清 et al. "中国贫困治理经验总结：扶贫政策能够实现有效增收吗？". In: 管理世界 38.2 (2022), pp. 70–83.

[59] Leonie Bossemeyer. "Machine Learning for Causal Discovery with Applications in Economics". MA thesis. Ludwig-Maximilians-Universität München, 2021.

[60] 韦颖璐. "基于 pair–copula 熵的相关性度量". 硕士学位论文. 苏州大学, 2021.

社会科学 II

- 管理学
  - 商品期货价格预测[61]
    - 国家统计局猪肉价格数据和大连商品交易所大豆期货价格数据
  - 单周期库存管理[62]
    - 大众朗逸汽车销售数据
- 社会学
  - 分析教育、职业和收入上的性别不平等问题[63]
    - 美国国家成年人收入调查数据（1994 年）
- 教育学
  - 高中数学成绩与其他学科成绩相关性分析[64]
    - 某市 2013 级理科学生高一、高二期末成绩和高三两次模考成绩

---

[61] Wuyue An, Lin Wang, and Dongfeng Zhang. "Comprehensive commodity price forecasting framework using text mining methods". In: *Journal of Forecasting* (2023). DOI: https://doi.org/10.1002/for.2985.

[62] Yu-Xin Tian and Chuan Zhang. "An end-to-end deep learning model for solving data-driven newsvendor problem with accessibility to textual review data". In: *International Journal of Production Economics* (2023), p. 109016. ISSN: 0925-5273. DOI: 10.1016/j.ijpe.2023.109016.

[63] Jian Ma. "Causal Domain Adaptation with Copula Entropy based Conditional Independence Test". In: *arXiv preprint arXiv:2202.13482* (2022). arXiv: 2202.13482 [cs.LG].

[64] 柳琼. "基于 Copula 和 MI 理论的相关性度量及其应用研究". 硕士学位论文. 三峡大学, 2018.

社会科学 III

- 计算语言学
  - 城市热线派单系统知识图谱构建[65]
    - 济南市民热线数据
- 新闻传播学
  - 上海新冠疫情下的公众情绪变化[66]
    - 微博平台“上海疫情”主题数据
- 法学
  - 社区属性与社区犯罪关系分析[67]
    - 美国社区与犯罪数据集

[65] 陈作海, 钱恒, and 高永超. “一种基于知识图谱的城市热线派单方法及系统”. Pat. CN115860436A. CN115860436A. 2023.

[66] Bowen Zhang et al. “Changes in Public Sentiment under the Background of Major Emergencies – Taking the Shanghai Epidemic as an Example”. In: *International Journal of Environmental Research and Public Health* 19.19 (2022), p. 12594. ISSN: 1660-4601. DOI: 10.3390/ijerph191912594. URL: https://www.mdpi.com/1660-4601/19/19/12594.

[67] Mario Wieser. “Learning Invariant Representations for Deep Latent Variable Models”. PhD thesis. University of Basel, 2020.

社会科学 IV

- 政治学
  - 分析政权领导力因素和政权危机之间关系[68]
    - 雪城大学莫伊尼汉全球事务研究所国际政治领导力数据集
- 军事学
  - 目标意图识别[69]
    - 空中飞行目标示例
- 情报学
  - 颠覆性技术科学-技术-产业互动模式分析[70]
    - 再生医学（干细胞）和白血病治疗相关资料数据

[68] Stuart William Card. "Towards an Information Theoretic Framework for Evolutionary Learning". MA thesis. Syracuse University, 2011.

[69] 张可 et al. "一种基于动态贝叶斯网络的目标意图识别方法". Pat. CN114997306A. CN114997306A. 2022.

[70] 许海云 et al. "颠覆性技术的科学-技术-产业互动模式识别与分析". In: 情报学报 42.7 (2023), pp. 816–831.

工程学 I

- 能源工程
  - 能源网络管理，研究天气因素与能源网络的耦合[71]
    - 北方某地区能源系统运行数据
  - 光伏发电功率预测[72]
    - 澳大利亚 Yulara 地区光伏电站数据
  - 风电机组工况划分[73]
    - 广东某海上风电场 SCADA 数据
  - 电力负荷预测[74]
    - 摩洛哥缔头万城电力消费数据
  - 风光储协同规划[75]
    - 某工业园区风光火储联合发电系统

[71] Xueqian Fu et al. "Uncertainty analysis of an integrated energy system based on information theory". In: *Energy* 122.122 (2017), pp. 649–662.

[72] 朱正林 and 张冕. "基于 AO 优化 VMD-CE-BiGRU 的光伏发电功率预测". In: 国外电子测量技术 41.10 (2022), pp. 56–61.

[73] 崔双双 and 孙单勋. "分工况下风电机组各变量相关性研究". In: 综合智慧能源 44.12 (2022), pp. 49–55.

[74] Jian Ma. "Identifying Time Lag in Dynamical Systems with Copula Entropy based Transfer Entropy". In: *arXiv preprint arXiv:2301.06037*, arXiv:2301.06037 (2023). arXiv: 2301.06037 [cs.LG].

[75] 董海艳 et al. "一种含源荷时序相似度约束的源储协同规划配置方法". Pat. CN114421538A. CN110766314A. 2022.

工程学 II

- 能源工程
  - 电网频率稳定性预测[76]
    - 贵州电网数据
  - 用户线损贡献分析[77]
    - 辽宁电网数据
  - 电价预测[78]
    - 2017 年美国 PJM 电力市场电价数据
  - 锂电池容量估计[79]
    - NASA 锂电池退化数据
  - 电力系统宽频振荡影响因素和传播路径分析[80]

---

[76] Peili Liu et al. "Frequency Stability Prediction of Power Systems Using Vision Transformer and Copula Entropy". In: *Entropy* 24.8 (2022), p. 1165. ISSN: 1099-4300. DOI: 10.3390/e24081165. URL: https://www.mdpi.com/1099-4300/24/8/1165.

[77] Wei Hu et al. "Research on User Loss Contribution Calculation of High-Loss Distribution Area Based on Transfer Entropy". In: *2022 China International Conference on Electricity Distribution (CICED)*. 2022, pp. 499–502. DOI: 10.1109/CICED56215.2022.9929052.

[78] Xiaoping Xiong and Guohua Qing. "A hybrid day-ahead electricity price forecasting framework based on time series". In: *Energy* (2022), p. 126099. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2022.126099. URL: https://www.sciencedirect.com/science/article/pii/S0360544222029851.

[79] Jiabei He and Lifeng Wu. "Cross-conditions capacity estimation of lithium-ion battery with constrained adversarial domain adaptation". In: *Energy* 277 (2023), p. 127559. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2023.127559. URL: https://www.sciencedirect.com/science/article/pii/S0360544223009532.

[80] 冯双 et al. "一种电力系统宽频振荡影响因素和传播路径分析方法". Pat. CN114977222A. CN114977222A. 2022.

工程学 III

- 食品工程
  - 葡萄酒质量与理化成分关系分析[81]
    - 葡萄牙绿酒葡萄酒理化成分与质量评价数据
- 土木建筑
  - 建筑能源系统节能技术[82]
    - 大连某教学楼供热监测数据
  - 工程变形监测[83]
    - 某隧道工程施工段围堰监测数据

---

[81] Marvin Lasserre, Régis Lebrun, and Pierre-Henri Wuillemin. "Learning Continuous High-Dimensional Models using Mutual Information and Copula Bayesian Networks". In: *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*. AAAI Press, 2021, pp. 12139–12146. URL: https://ojs.aaai.org/index.php/AAAI/article/view/17441, Marvin Lasserre. "Apprentissages dans les réseaux bayésiens à base de copules non-paramétriques". PhD thesis. Sorbonne Université, 2022. URL: https://tel.archives-ouvertes.fr/tel-03647090.

[82] Zhiwei Li et al. "A model-free method for identifying time-delay characteristics of HVAC system based on multivariate transfer entropy". In: *Building and Environment* 217 (2022), p. 109072. ISSN: 0360-1323. DOI: https://doi.org/10.1016/j.buildenv.2022.109072. URL: https://www.sciencedirect.com/science/article/pii/S0360132322003110.

[83] 曹久慧 et al. "一种基于自注意力机制的变形监测方法". Pat. CN116378120A. CN116378120A. 2023.

工程学 IV

- 交通工程
  - 大件货物运输方案制定[84]
    - 大件货物运输案例数据
  - 航空和高铁票价影响因素分析[85]
    - 京沪高铁和航空票价数据
  - 城市轨道交通客流分析和预测[86]
    - 苏州市轨道交通系统客流时序数据

[84] 黄达. "基于模块链构建的大件货物多式联运方案研究". 博士学位论文. 北京交通大学, 2021.

[85] 许罗豪 et al. "基于熵与回归树的票价影响因素研究". In: 综合运输 45.6 (2023), pp. 125–130.

[86] 王升. "基于多源数据的城市轨道交通系统客流分析与预测". 硕士学位论文. 东南大学, 2022.

工程学 V

- 制造工程
  - 制造质量管理，研究优化制造过程参数，预测产品质量[87]
    - 富士康生产线制造过程数据
  - 装配质量控制[88]
    - 江淮汽车某型汽油发动机关键零部件装配过程数据
  - 工业过程故障监测[89]
    - 鞍钢热轧带钢工艺过程数据
  - 钢铁工艺过程碳排放预测[90]
    - 某钢铁厂烧结过程数据
  - 液晶显示器质量预测[91]
    - 薄膜晶体管液晶显示器生产数据

---

[87] Yan-Ning Sun et al. "Modelling and Prediction of Injection Molding Process Using Copula Entropy and Multi-Output SVR". In: *IEEE 17th International Conference on Automation Science and Engineering*. 2021.

[88] 王小巧. "复杂机械产品装配过程质量自适应控制方法及支持系统研究". 博士学位论文. 合肥工业大学, 2015.

[89] Jie Dong, Keren Cao, and Kaixiang Peng. "Hierarchical Causal Graph-Based Fault Root Cause Diagnosis and Propagation Path Identification for Complex Industrial Process Monitoring". In: *IEEE Transactions on Instrumentation and Measurement* 72 (2023), pp. 1–11. DOI: 10.1109/TIM.2023.3268464.

[90] Jie Hu et al. "Dynamic Modeling Framework Based on Automatic Identification of Operating Conditions for Sintering Carbon Consumption Prediction". In: *IEEE Transactions on Industrial Electronics* (2023), pp. 1–9. DOI: 10.1109/TIE.2023.3270514.

[91] Hongxia Cai and Zhiqiang Rong. "Key Quality Feature Identification and Quality Prediction in Complex Manufacturing Processes". In: *2023 15th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. 2023, pp. 229–232. DOI: 10.1109/IHMSC58761.2023.00060.

工程学 VI

- 可靠性工程
  - 系统退化过程建模[92]
    - 微波电子组件数据
  - 风电机组健康状态评估[93]
    - 内蒙古某风场的风机 SCADA 数据
- 化学工程
  - 化学过程故障监测和诊断[94]
    - Tennessee Eastman 过程数据
  - 化工过程因果网络构建[95]
    - 连续搅拌槽式反应器数据和 Tennessee Eastman 过程数据

---

[92] Fuqiang Sun et al. "A Copula Entropy Approach to Dependence Measurement for Multiple Degradation Processes". In: *Entropy* 21.8 (2019), p. 724.

[93] 齐咏生 et al. "一种基于多维度 SCADA 数据评估风电机组健康状态评估方法". Pat. CN110442833A. CN110442833A. 2019.

[94] Min Yin, Jince Li, and Hongguang Li. "A CNN approach based on correlation metrics to chemical process fault classifications with limited labeled data". In: *The Canadian Journal of Chemical Engineering* (2022). DOI: 10.1002/cjce.24749. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cjce.24749. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cjce.24749, Yingpeng Wei and Li Wang. "Copula entropy-based PCA method and application in process monitoring". In: *2022 4th International Conference on Intelligent Information Processing (IIP)*. 2022, pp. 61–64. DOI: 10.1109/IIP57348.2022.00019.

[95] Xiaotian Bi et al. "Large-scale chemical process causal discovery from big data with Transformer-based deep learning". In: *Process Safety and Environmental Protection* (2023).

工程学 VII

- 航空航天
  - 飞行器总体参数分析和优化[96]
    - 美国喷气战斗机总体设计参数数据
  - 卫星在轨健康状态监测[97]
    - 真实卫星遥测数据
    - NASA 公开的 SMAP 和 MSL 数据集
  - 涡扇发动机健康状态监测[98]
    - NASA 格林中心引擎性能退化模拟数据
  - 机场间航班延误因果关系分析[99]
- 车辆工程
  - CAN 总线入侵检测[100]
    - 现代汽车 YF 索纳塔 CAN 数据

[96] Baby Alpettiyil Krishnankutty, Rajesh Ganapathy, and Paduthol Godan Sankaran. "Non-parametric estimation of copula based mutual information". In: *Communications in Statistics - Theory and Methods* 49.6 (2020), pp. 1513–1527. DOI: 10.1080/03610926.2018.1563180.

[97] Hao Liu et al. "Data-driven identification model for associated fault propagation path". In: *Measurement* 188 (2022), p. 110628. ISSN: 0263-2241. DOI: 10.1016/j.measurement.2021.110628, Zefan Zeng et al. "Satellite Telemetry Data Anomaly Detection Using Causal Network and Feature-Attention-Based LSTM". In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–21. DOI: 10.1109/TIM.2022.3151930.

[98] 贾如侠. "涡扇发动机故障预测及剩余寿命分析方法研究". 硕士学位论文. 哈尔滨师范大学, 2023.

[99] 吴格 et al. "一种因果关系分析方法及装置". Pat. CN110766314A. CN110766314A. 2020.

[100] Sheng Gao et al. "Attack Detection for Intelligent Vehicles via CAN- Bus: A Lightweight Image Network Approach". In: *IEEE Transactions on Vehicular Technology* (2023), pp. 1–13. DOI: 10.1109/TVT.2023.3296705.

工程学 VIII

- 电子工程
  - 集成电路封装材料物理性能预测[101]
    - CuNi 合金体系材料强度和稳定性计算数据
- 通信工程
  - 通讯网络加密技术研究[102]
  - 6G 网络语义通信技术研究[103]
    - ImageNet-1k 数据集和 VOC2012 数据集
- 高性能计算
  - 高性能计算能源效率优化[104]
- 测绘遥感
  - 高光谱遥感数据分析[105]
    - 美国印第安纳 Indian Pine 高光谱遥感数据

---

[101] 刘勃. "基于机器学习的封装材料加速预测". 硕士学位论文. 哈尔滨理工大学, 2022.

[102] Xu Wang et al. "Physical Layer Secret Key Capacity Using Correlated Wireless Channel Samples". In: *2016 IEEE Global Communications Conference (GLOBECOM)*. 2016, pp. 1–6.

[103] 傅宇舟 et al. "面向 6G 网络的基于语义通信的端到端服务框架". In: 移动通信 47.6 (2023), pp. 35–40.

[104] Andreas Gocht-Zech. "Ein Framework zur Optimierung der Energieeffizienz von HPC-Anwendungen auf der Basis von Machine-Learning-Methoden". PhD thesis. Technische Universität Dresden, 2022.

[105] Xuexing Zeng and T S Durrani. "Band selection for hyperspectral images using copulas-based mutual information". In: *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*. 2009, pp. 341–344. DOI: 10.1109/SSP.2009.5278570.

金融工程 I

- 投资组合优化
  - 股票资产相关性网络分析[106]
    - 沪深 A 股指数、沪深 300 指数数据
  - ST 股票分类[107]
    - A 股市场 ST 股票数据
- 金融问题建模
  - Copula 函数模型选择[108]
    - 标普 500 指数数据
- 股票相关性建模
  - R-vine copula 结构建模[109]
    - 德国 DAX 指数数据
    - 中证五大行业指数数据

---

106 Qiutong Wang. "Social Networks, Asset Allocation and Portfolio Diversification". MA thesis. University of Waterloo, 2015.

107 朱仲儿. "多种机器学习方法的股票分类预测". 硕士学位论文. 上海师范大学, 2022.

108 Rafael Calsaverini and Renato Vicente. "An information-theoretic approach to statistical dependence: Copula information". In: *EPL (Europhysics Letters)* 88.6 (2009), p. 68003. DOI: 10.1209/0295-5075/88/68003.

109 Fadhah Amer Alanazi. "Truncating Regular Vine Copula Based on Mutual Information: An Efficient Parsimonious Model for High-Dimensional Data". In: *Mathematical Problems in Engineering* 2021 (2021), p. 4347957, 王念鸽. "基于互信息的 Vine Copula 模型的高频数据投资组合风险测度研究". 硕士学位论文. 浙江财经大学, 2023.

金融工程 II

- 量化金融工具箱 MLFinLab[110]
- 金融系统性风险
  - 行业风险溢出效应分析[111]
    - 我国股票市场 11 个行业交易数据
  - 金融脆弱性度量[112]
    - 沪深 300 指数股票数据
- 信用风险评价
  - 信用风险卡模型建立[113]
    - 信用卡客户数据
- 金融产品价格预测
  - 基于因果关系的迁移学习价格预测模型构建[114]
    - 国际主要金融指数、能源期货价格和农产品价格数据

---

[110] Hudson and Thames. *Machine Learning Financial Laboratory (MLFinLab)*. GitHub. 2021. URL: https://github.com/hudson-and-thames/mlfinlab.

[111] 熊靖宇. "基于 Copula 熵的行业风险溢出效应分析". 硕士学位论文. 东北财经大学, 2020.

[112] Mengyuan Chen et al. "Vulnerability Analysis Method Based on Network and Copula Entropy". In: *Preprints* (2023).

[113] 孔祥永 et al. "一种自动化特征工程信用风险评价系统及方法". Pat. CN114049198A. CN114049198A. 2021.

[114] Dabin Zhang et al. "A novel deep transfer learning framework with adversarial domain adaptation: application to financial time-series forecasting". In: *Neural Computing and Applications* (2023). URL: https://doi.org/10.1007/s00521-023-09047-1.

My Golf

Enjoy the Power of Copula Entropy!