Copula熵的多学科应用

2021-12-29

综述论文

- 马健. Copula熵：理论和应用. ChinaXiv:202105.00070 (2021).

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性表示的理论工具，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，包括其在统计学四个基本问题（结构学习、关联发现、变量选择和时序因果发现等）上的理论应用。讨论了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的相关性度量框架进行了对比。简述了Copula熵在水文学、环境气象学、生态学、化学信息学、认知神经学、计算神经学、系统生物学、生物信息学、临床诊断学、老年医学、公共卫生学、经济政策学、政治学，以及能源工程、制造工程、可靠性工程、航空工程、通信工程、测绘工程和金融工程等领域的实际应用。

水文学

- 洪水预报
  - 金沙江流域洪水预报 [Chen2013, Chen2014]
- 水文事件风险建模
  - 干旱事件识别 [Ni2020]
- 水文观测网络选址和优化
  - 上海降水观测网 [Xu2017]
  - 汾河流域观测网 [Li2020]
  - 北京市区水文观测网 [Li2020]
  - 太湖盆地流域降水观测网 [Li2020]
- 多站点径流生成
  - 巴西雅瓜拉比-大都市水库系统 [Porto2021]
- 中长期径流预报
  - 南水北调工程丹江口水库入库径流预报 [黄2021]

环境气象学

- 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[Ma2019a]
    - 北京地区PM2.5和气象观测数据

生态学

- 动物运动轨迹分析
  - Cylcop算法包 [Hodel2021]

化学信息学

- 分子设计
  - 设计具有特定属性的分子结构 [Wieser2020]
    - 有机分子属性QM9数据库

神经科学

- 认知神经学
  - 分析大脑认知活动的多模态数据 [Kayser2015, Ince2016, Ince2017]
    - 人脸检测任务EEG数据和听觉语音刺激任务MEG数据
- 计算神经学
  - 神经元可塑性建模 [Leugering2018]
  - 神经网络信息传输关系分析 [Parman2021]

生物学

- 系统生物学
  - 生物信号调控和传导 [Charzynska2015]
    - 癌症分子机制数据
  - 生物现象动态网络结构和功能 [Farhangmehr2013]
    - 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 [Wieczorek2016]
    - 肝炎病毒感染治疗基因表达谱数据

医学

- 临床医学
  - 心脏病诊断 [Ma2021]
    - UCI心脏病数据
  - 糖尿病病情管理 [Mesiar2021]
    - 美国Health Facts糖尿病救治网络数据
- 认知医学
  - 认知能力评估／痴呆症筛查 [Ma2019b]
    - 北京、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 [Ma2020a, Ma2020b]
    - 天津、成都跌倒人群老年人数据

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 [Mesiar2021]
    - 新冠临床数据

社会科学

- 经济政策学
  - 扶贫政策效果评估，用于政策目标人口鉴别 [Shan2020]
    - 2018年政府贫困家庭状况普查数据（四川省）
- 政治学
  - 分析政权领导力因素和政权危机之间关系 [Card2011]
    - 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

工程

- 能源工程
  - 能源网络管理，研究天气因素与能源网络的耦合 [Fu2017]
    - 北方某地区能源系统运行数据
- 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 [Sun2021]
    - 富士康生产线制造过程数据
- 可靠性工程
  - 系统退化过程建模 [Sun2019]
- 航空工程
  - 飞行器总体参数分析和优化 [Krishnankutty2020]
    - 美国喷气战斗机总体设计参数数据
- 通信工程
  - 通讯网络加密技术研究 [Wang2016]
- 测绘工程
  - 高光谱遥感数据分析 [Zeng2009]
    - 美国印第安纳Indian Pine高光谱遥感数据

金融工程

- 量化金融工具箱 **MLFinLab**
  - Hudson and Thames Quantitative Research [HudsonThames2021]
    - 非线性相关分析算法
- 投资组合优化
  - 股票资产相关性网络分析 [Wang2015]
    - 沪深A股指数、沪深300指数数据
- 金融问题建模
  - Copula函数模型选择 [Calsaverini2009]
    - 标普500指数数据

参考文献

- [Chen2013] Lu Chen, Vijay P. Singh, and Shenglian Guo. Measure of correlation between river flows using the copula-entropy method. Journal of Hydrologic Engineering, 18(12):1591–1606, 2013.

- [Chen2014] Lu Chen, Vijay P. Singh, Shenglian Guo, Jianzhong Zhou, and Lei Ye. Copula entropy coupled with artificial neural network for rainfall–runoff simulation. Stochastic Environmental Research and Risk Assessment, 28(7):1755–1767, 2014.

- [Ni2020] Lingling Ni, Dong Wang, Jianfeng Wu, Yuankun Wang, Yuwei Tao, Jianyun Zhang, Jiufu Liu, and Fei Xie. Vine copula selection using mutual information for hydrological dependence modeling. Environmental Research, 186:109604, 2020.

- [Xu2017] Pengcheng Xu, Dong Wang, Vijay P. Singh, Yuankun Wang, Jichun Wu, Lachun Wang, Xinqing Zou, Yuanfang Chen, Xi Chen, Jiufu Liu, Ying Zou, and Ruimin He. A two-phase copula entropy-based multiobjective optimization approach to hydrometeorological gauge network design. Journal of Hydrology, 555:228–241, 2017.

- [Li2020] Heshu Li, Dong Wang, Vijay P. Singh, Yuankun Wang, Jianfeng Wu, Jichun Wu, Ruimin He, Ying Zou, Jiufu Liu, and Jianyun Zhang. Developing a dual entropy-transinformation criterion for hydrometric network optimization based on information theory and copulas. Environmental Research, 180:108813, 2020.

- [Porto2021] Victor Costa Porto, Francisco de Assis de Souza Filho, Taís Maria Nunes Carvalho, Ticiana Marinho de Carvalho Studart, and Maria Manuela Portela. A GLM copula approach for multisite annual streamflow generation. Journal of Hydrology, 598:126226, 2021.

- [黄2021] 黄朝君, 贾建伟, 秦赫, 王栋. 基于Copula熵 - 随机森林的中长期径流预报研究. 人民长江, 2021, 52(11): 81-85.

- [Ma2019a] Jian Ma. Estimating transfer entropy via copula entropy. arXiv preprint arXiv:1910.04375, 2019.

- [Hodel2021] Florian H. Hodel and John R. Fieberg. Cylcop: An R package for circularlinear copulae with angular symmetry. bioRxiv, page 2021.07.14.452253, 2021.

- [Wieser2020] Mario Wieser, Sonali Parbhoo, Aleksander Wieczorek, and Volker Roth. Inverse learning of symmetries. In Advances in Neural Information Processing Systems, volume 33, pages 18004–18015, 2020.

- [Ince2017] Robin A.A. Ince, Bruno L. Giordano, Christoph Kayser, Guillaume A. Rousselet, Joachim Gross, and Philippe G. Schyns. A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. Human Brain Mapping, 38(3):1541–1573, 2017.

- [Ince2016] Robin A. A. Ince, Katarzyna Jaworska, Joachim Gross, Stefano Panzeri, Nicola J. van Rijsbergen, Guillaume A. Rousselet, and Philippe G. Schyns. The deceptively simple N170 reflects network information processing mechanisms involving visual feature coding and transfer across hemispheres. Cerebral Cortex, 26(11):4123–4135, 2016.

- [Kayser2015] Stephanie J. Kayser, Robin A.A. Ince, Joachim Gross, and Christoph Kayser. Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. The Journal of Neuroscience, 35(44):14691–14701, 2015.

- [Leugering2018] Johannes Leugering and Gordon Pipa. A unifying framework of synaptic and intrinsic plasticity in neural populations. Neural Computation, 30(4):945–986, 2018.

- [Parman2021] Ari Pakman, Amin Nejatbakhsh, Dar Gilboa, Abdullah Makkeh, Luca Mazzucato, Michael Wibral, and Elad Schneidman. Estimating the unique information of continuous variables in recurrent networks. Advances in Neural Information Processing Systems, 2021.

- [Charzynska2015] Agata Charzyńska and Anna Gambin. Improvement of the k-NN entropy estimator with applications in systems biology. Entropy, 18(1):13, 2015.

参考文献 (cont)

- [Farhangmehr2013] Farzaneh Farhangmehr, Daniel M. Tartakovsky, Parastou Sadatmousavi, Mano R. Maurya, and Shankar Subramaniam. An information-theoretic algorithm to data-driven genetic pathway interaction network reconstruction of dynamic systems. In 2013 IEEE International Conference on Bioinformatics and Biomedicine, pages 214–217, 2013.

- [Wieczorek2016] Aleksander Wieczorek and Volker Roth. Causal compression. arXiv preprint arXiv:1611.00261, 2016.

- [Ma2021] Jian Ma. Variable selection with copula entropy. Chinese Journal of Applied Probability and Statistics, 37(4):405–420, 2021. See also arXiv preprint arXiv:1910.12389 (2019).

- [Mesiar2021] Radko Mesiar and Ayyub Sheikhi. Nonlinear random forest classification, a copula-based approach. Applied Sciences, 11(15), 2021.

- [Ma2019b] Jian Ma. Predicting MMSE score from finger-tapping measurement. bioRxiv, page 817338, 2019.

- [Ma2020a] Jian Ma. Predicting TUG score from gait characteristics based on video analysis and machine learning. bioRxiv, page 963686, 2020.

- [Ma2020b] Jian Ma. Associations between finger tapping, gait and fall risk with application to fall risk assessment. arXiv preprint arXiv:2006.16648, 2020.

- [Shan2020] Qingsong Shan and Qianning Liu. Binary trees for dependence structure. IEEE Access, 8:150989–150998, 2020.

- [Card2011] Stuart William Card. Towards an information theoretic framework for evolutionary learning. Master's thesis, Syracuse University, 2011.

- [Fu2017] Xueqian Fu, Hongbin Sun, Qinglai Guo, Zhaoguang Pan, Wen Xiong, and Li Wang. Uncertainty analysis of an integrated energy system based on information theory. Energy, 122(122):649–662, 2017.

- [Sun2021] Yan-Ning Sun, Yu Chen, Wu-Yin Wang, Hong-Wei Xu, and Wei Qin. 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.

- [Sun2019] Fuqiang Sun, Wendi Zhang, Ning Wang, and Wei Zhang. A copula entropy approach to dependence measurement for multiple degradation processes. Entropy, 21(8):724, 2019.

- [Krishnankutty2020] Baby Alpettiyil Krishnankutty, Rajesh Ganapathy, and Paduthol Godan Sankaran. Non-parametric estimation of copula based mutual information. Communications in Statistics - Theory and Methods, 49(6):1513–1527, 2020.

- [Wang2016] Xu Wang, Liang Jin, Kaizhi Huang, Mingliang Li, and Yi Ming. Physical layer secret key capacity using correlated wireless channel samples. In 2016 IEEE Global Communications Conference (GLOBECOM), pages 1–6, 2016.

- [Zeng2009] 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, pages 341–344, 2009.

- [HudsonThames2021] Hudson and Thames. Machine learning financial laboratory (MLFinLab), 2021. URL: https://github.com/hudson-and-thames/mlfinlab.

- [Wang2015] Qiutong Wang. Social networks, asset allocation and portfolio diversification. Master's thesis, University of Waterloo, 2015.

- [Calsaverini2009] Rafael Calsaverini and Renato Vicente. An information-theoretic approach to statistical dependence: Copula information. EPL (Europhysics Letters), 88(6):68003, 2009.

谢谢 ！

欢迎提出宝贵意见和建议