



面向世界科技前沿, 面向国家重大需求, 面向国民经济主战场, 率先实现科学技术跨越发展, 率先建成国家创新人才高地, 率先建成国家高水平科技智库, 率先建设国际一流科研机构。

中国科学院办院方针



搜索

重庆研究院半监督分类学习技术研究获进展

文章来源: 重庆绿色智能技术研究院 发布时间: 2017-09-21 【字号: 小 中 大】

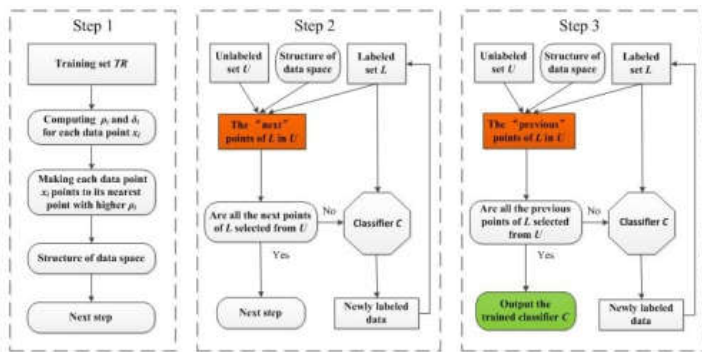
我要分享

近日, 中国科学院重庆绿色智能技术研究院大数据挖掘及应用中心团队对半监督分类学习及其应用开展的研究, 取得了系列进展。相关研究成果发表在IEEE Transactions on Industrial Informatic、Neurocomputing和Ecological Indicators等期刊上, 研究获得国家科技重大专项“水体污染防治与治理”项目、中科院率先行动百人计划项目和国家自然科学基金项目的支持。

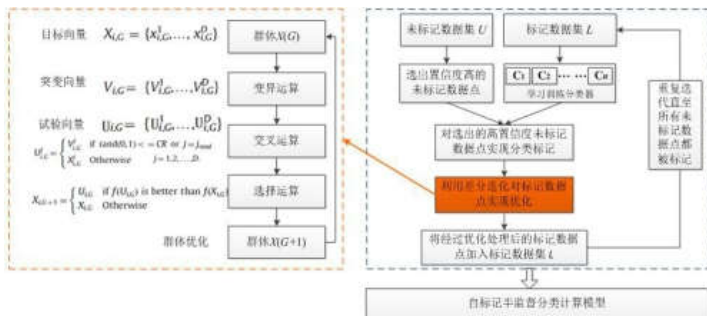
监督学习, 是两种基本的机器学习方式之一, 其基本原理是通过分析大量已标记的数据信息, 根据某些假设挖掘出数据信息的内部规律, 建立相应学习系统, 从而对未见样本信息进行估计与预测。然而, 在现实生产生活中获取足够多的标记数据较为困难, 这需要耗费大量的人力物力实现对海量数据的预先标记。在大数据时代, 海量未标记数据与少量标记数据同时存在, 如何通过仅有的少量标记数据和海量未标记数据, 来学习实现优秀分类系统显得尤为重要。同时, 如何利用未标记数据辅助少量标记数据进行学习, 以改进分类学习系统性能, 近年来成为机器学习领域研究热点, 并被广泛应用于工业、农业和交通等领域。

研究团队针对现有传统自标记半监督分类模型的局部最优解、仅适用球状数据问题, 提出了一种基于数据密度峰值的自标记半监督分类模型。该模型通过无监督的计算数据密度峰值, 发现数据内部聚类知识空间结构特征, 将此聚类知识空间结构特征引入自标记半监督分类学习训练过程, 可实现高质量的半监督分类计算。科研人员针对传统自标记半监督分类模型在自训练迭代过程中的错误标记问题, 提出了一种适应于所有自标记半监督分类模型的优化框架。该优化框架通过将差分进化算法引入自标记半监督分类模型的自训练迭代过程, 实现对所有标记数据点的监督优化, 可最大程度避免错误标记问题在自训练迭代过程中出现, 提升分类学习系统性能。此外, 研究团队针对水体富营养化这一全球性水环境问题, 提出了一种基于自标记半监督分类算法的水体富营养化评价模型。该评价模型通过将上述半监督分类学习应用于水体富营养化评价, 解决了传统富营养化评价方法与模型面临的所需指标获取代价太高和实时监测大数据处理能力不足的问题, 研究成果将有助于在大数据时代下对水库富营养化问题实现科学认知与评价。

论文链接: 1 2 3



基于数据密度峰值的自标记半监督分类模型流程图



自标记半监督分类模型优化框架示意图

热点新闻

中国科大建校60周年纪念大会举行

- 中科院召开党建工作推进会
驻中科院纪检监察组发送中秋国庆期间廉...
中科院党组学习贯彻习近平总书记在全国...
国科大举行2018级新生开学典礼
中科院党组学习研讨药物研发和集成电路...

视频推荐

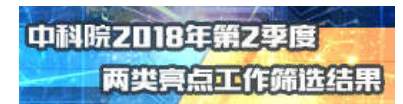


【新闻联播】“率先行动”计划 领跑科技体制改革



【新闻直播间】中科院2018年第三季度新闻发布会: “丝路环境”专项昨日正式启动

专题推荐



(责任编辑: 侯茜)



© 1996 - 2018 中国科学院 版权所有 京ICP备05002857号 京公网安备110402500047号 联系我们
地址: 北京市三里河路52号 邮编: 100864