

◆生物医学信息计算◆

基于 BTLBOGSA 与 CNN 的基因微阵列数据分类模型*

赵晚昭, 谢 聪**

(广西农业职业技术大学, 广西南宁 530007)

摘要:针对现有基因微阵列数据分类中存在的维度高、容易发生过拟合的问题,提出了基于 BTLBOGSA (Binary TLBOGSA) 与卷积神经网络 (Convolutional Neural Network, CNN) 的基因微阵列数据分类模型 (BTLBOGSA-CNN)。该模型首先针对基因微阵列数据分类时存在的维度高的问题,利用新的编码策略,将连续搜索空间转换为二元搜索空间,结合教与学优化 (Teaching-Learning-Based Optimization, TLBO) 算法的二元变体与引力搜索算法 (Gravitational Search Algorithm, GSA) 的各自特点,基于 BTLBOGSA 方法从基因微阵列数据集中选择具有高鉴别性的基因;然后针对基因微阵列数据分类易发生过拟合问题的现象,利用卷积神经网络进行基因微阵列数据的分类。利用公开的基因微阵列数据集进行仿真实验,从 TLBO 算法与 GSA 结合的有效性、BTLBOGSA 与 CNN 结合的有效性、BTLBOGSA-CNN 与其他已有分类模型相比的有效性 3 个方面进行对比分析,结果表明,BTLBOGSA-CNN 模型可以在较少的特征基因下取得较高的分类精度,具有较高的可行性和有效性。

关键词:基因微阵列数据分类 教与学优化算法 引力搜索算法 特征选择 卷积神经网络

中图分类号: TP301 文献标识码: A 文章编号: 1005-9164(2022)02-0260-09

DOI: 10.13656/j.cnki.gxkx.20220526.005

近年来,癌症及恶性肿瘤一直威胁着人类的生命健康,尤其是患病早期症状不容易被查出,导致错过了最佳的治疗时间,因此,为各种疾病提供一种高效、准确的诊断方法,不仅可以使患者及时接受治疗,或许还可以挽救患者生命。随着基因微阵列技术的成熟,基因表达谱可以表示人类各个组织的正常基因,当前很多重大疾病的基因微阵列数据也已被共享,可

为基因分类与识别提供大量可靠的数据。一般来说,在高维微阵列数据中,冗余的基因不仅会降低训练强度,而且会对学习算法的性能产生负面的影响。为了解决这些问题,研究人员提出了众多基因选择方法来选择最佳的鉴别基因^[1]。

在进行基因微阵列分类时,最重要的是对微阵列基因进行特征选择,当前的基因选择方法主要分为过

收稿日期: 2021-11-19

* 广西自然科学基金项目(2021GXNSFBA220080),广西高等教育本科教学改革工程项目(2021JGA425),广西中青年骨干教师科研基础能力提升项目(2021KY1736)和国家自然科学基金企业创新发展联合基金(U19B2021)资助。

【作者简介】

赵晚昭(1985-),男,高级实验师,主要从事智能化应用和计算机控制技术研究。

【**通信作者】

谢 聪(1982-),男,教授/高级工程师,主要从事智能算法和网络安全研究,E-mail:wplhns@163.com。

【引用本文】

赵晚昭,谢聪.基于 BTLBOGSA 与 CNN 的基因微阵列数据分类模型[J].广西科学,2022,29(2):260-268.

ZHAO W Z, XIE C. Classification Model of Gene Microarray Data Based on BTLBOGSA and CNN [J]. Guangxi Sciences, 2022, 29(2): 260-268.