

基于证据理论的数据分类研究综述

黄林庆¹, 范金福², 刘功申¹, Alan Wee-Chung Liew³

¹上海交通大学计算机学院, 中国上海市, 200240

²青岛大学计算机科学技术学院, 中国青岛市, 266071

³格里菲斯大学信息与通信技术学院, 澳大利亚昆士兰州, 4215

摘要: 数据分类是机器学习与数据分析领域的基础任务, 应用场景覆盖诸多行业。在实际应用中, 特征区分度低、缺失值、分布偏移、类别不平衡、标签空间不一致、类别重叠等因素都会带来不确定性, 而依托概率框架构建的传统分类方法难以显性表达并弱化这类不确定性。证据理论(ET, 亦称作Dempster-Shafer理论)通过基本信任分配与证据组合规则, 搭建起可对不确定性和模糊性建模的灵活框架, 在分类领域受到日益广泛的重视, 基于证据理论的数据分类(DCET)随之成为热门研究课题。本文系统性梳理DCET, 按照基于特征和基于标签两类不确定性场景对现有方法进行归类。首先, 介绍证据理论基础原理, 总结证据 K 近邻分类器及其衍生改进算法; 其次, 分别梳理高维特征、属性缺失、特征分布偏移、标签分布失衡、标签空间不一致、类别重叠六大不确定性场景下适用于表格数据分类的证据理论方法; 继而, 阐述证据理论与深度神经网络结合、面向图像这类非结构化数据的分类方案; 此外, 归纳了人体行为识别、医学图像分割、遥感影像分类、遥感影像变化检测等代表性应用场景; 最后, 探讨了DCET未来研究方向。

关键词: 证据理论; 不确定数据分类; 信任函数; 证据推理; D-S理论; 多源多源信息融合

<https://doi.org/10.1631/ENG.ITEE.2026.0125>