

基于机器学习的资产评估模型构建与应用研究

韩萌

陕西服装工程学院, 中国·陕西 西安 710021

摘要: 资产评估模型的构建与应用研究, 特别是基于机器学习的方式, 正成为金融和经济领域的热门话题。论文讨论了机器学习技术在资产评估中的应用原则, 分析了构建此类模型时所面临的难点与挑战, 提出了有效的策略以加强机器学习在资产评估中的应用, 旨在为相关从业者和研究人员提供新的思路, 以更好地利用机器学习技术来提升资产评估的准确性和效率。

关键词: 机器学习; 资产评估; 模型构建; 应用; 策略

Research on the Construction and Application of Asset Evaluation Model Based on Machine Learning

Meng Han

Shaanxi Institute of Fashion Engineering, Xi'an, Shaanxi, 710021, China

Abstract: The construction and application research of asset valuation models, especially machine learning based methods, is becoming a hot topic in the fields of finance and economics. This paper discusses the application principles of machine learning technology in asset evaluation, analyzes the difficulties and challenges faced in building such models, and proposes effective strategies to strengthen the application of machine learning in asset evaluation. The aim is to provide new ideas for relevant practitioners and researchers to better utilize machine learning technology to improve the accuracy and efficiency of asset evaluation.

Keywords: machine learning; asset valuation; model construction; application; strategy

0 前言

资产评估在经济和金融决策中起着关键作用, 准确的评估对资本市场的稳定与资源有效配置至关重要。传统资产评估方法依赖人工经验和规则模型, 容易受到主观因素影响, 难以应对市场的复杂性和多变性。机器学习作为一种数据驱动的智能技术, 能够自动识别数据中的潜在规律和模式, 从而提供更精准和科学的评估方法。将机器学习应用于资产评估不仅能提升评估的效率和可靠性, 还推动资产管理和金融服务的智能化与数字化, 因此具备重要的现实意义和研究价值。

1 机器学习技术应用于资产评估模型构建的原则

1.1 数据质量原则

数据质量在构建机器学习资产评估模型时显得尤为关键, 数据来源必须可靠, 采集过程也要具备准确性和可信性, 不准确或有疑问的数据可能引发模型偏差, 对评估结果产生不利影响, 数据需保持无误差状态, 以准确反映真实信息, 因此在数据预处理阶段应仔细清洗和验证, 去除所有重复、错误或不一致的信息, 通过使用全面且多样的数据集, 可以更好地捕捉资产特性及市场动态, 避免样本不足引发模型过

拟合或泛化能力不足的问题。此外, 选择合适的数据来源和特征变量能够显著提升模型的预测表现, 为了确保模型性能, 需有效防止数据缺失, 缺失数据不仅会降低模型的准确性, 还可能带来偏差影响评估的客观性, 因此在数据收集和处理阶段应制定具体策略应对缺失数据, 如使用插补、删除或模型预测等方法。

1.2 特征选择原则

资产评估模型的构建中, 选对特征直接决定了模型的效果和结果是否容易理解, 必须深入理解资产的内在属性和市场行为, 才能找到真正与资产价值相关的特征, 确保这些特征可以准确反映价值变化, 如房地产评估时, 地理位置、建筑面积、房龄和周边设施等因素通常对市场价值有明显影响, 去除多余特征同样关键, 因为这些多余的特征不仅增加了模型复杂性, 还会导致模型在训练数据表现良好但在新数据上效果不佳, 可以通过主成分分析 (PCA) 或特征重要性评估等机器学习技术来识别并移除那些与其他特征高度相关或对预测贡献不大的特征。此外, 优先选择容易理解和解释的特征有助于提升模型的应用效果, 因为即使复杂模型在准确性上表现较好, 但内部机制难以解释会影响决策者对结果的信任, 因此选择那些能直观反映资产价值变化的特征不仅能提高可解释性, 还使分析和决策更为清晰, 用简单的经济指标或市场趋势作为特征, 可以帮助非专业人士更好地理

解预测依据。

1.3 模型选择原则

资产评估模型构建中,模型选择至关重要,选择合适的机器学习算法需深入理解数据的分布特性和潜在规律,因不同算法对数据的假设和处理方式各异,分析数据特性如分布形式、特征相关性和缺失值情况,能够帮助筛选出最适合的模型。随着数据规模不断扩大,模型的适应性显得尤为重要,优秀模型应能有效处理大规模数据,同时保持预测性能,这要求选择算法时需考虑计算效率和可扩展性。例如,某些模型在小数据集上表现良好,但在更大数据上可能出现过拟合或性能下降,因此选择在数据量增加时依然能稳定输出的模型,如集成学习算法或基于树的模型,通常是明智的选择。模型复杂度也是关注重点,尽管复杂模型在训练数据上表现优秀。但实际应用中,简化模型往往能提高可解释性和泛化能力,在资产评估领域,模型的透明性和可解释性尤其重要,决策者需理解模型依据,以便做出合理判断。

1.4 过拟合防范原则

机器学习技术用于构建资产评估模型时,过拟合问题往往会对模型效果产生负面影响,通常表现为模型在训练数据上的表现出色,但在新数据上的预测能力却较差,导致这一问题的主要原因是模型过于复杂,从而将数据中的噪声当成了有效模式。避免过拟合的有效方法之一是交叉验证,通过将数据集分成若干子集,让模型分别在不同子集上进行训练和测试,以充分评估模型在未见数据上的表现,同时正则化技术也被广泛应用以控制模型的复杂性,提升泛化能力。正则化通过引入惩罚项来限制模型参数的大小,使得模型在拟合训练数据时更加平滑,降低对噪声的敏感性。交叉验证和正则化的结合,不仅提升了模型的稳定性和可靠性,也增强了模型在实际市场评估中的应用效果,使其能够更准确地反映市场真实情况,交叉验证与正则化技术的应用有效防止过拟合,从而确保模型在不同情境下都能保持稳健和均衡的表现。

2 基于机器学习的资产评估模型构建的难点和挑战

2.1 数据质量和数据获取

构建基于机器学习的资产评估模型时,面临的关键挑战在于数据的质量和获取,资产评估需要大量的财务数据、市场数据及其他相关信息支撑,这些数据源涵盖了企业财务报表、市场行情以及宏观经济指标等内容,然而,实际获取的数据往往不够完整,存在缺失值、错误数据或者格式不一致等问题,数据的不完整性在模型训练中容易引入偏差,难以准确反映真实的资产价值,数据的准确性同样至关重要,金融数据牵涉多方面信息,不同机构和来源提供的资料可能存在矛盾,从而影响数据的可信度,如财务报表的审计质量、数据收集的方法和时间等因素都可能削弱数据的真实性,这些不一致性不但增加了数据处理的难度,也妨碍了模型的预

测效果,更为复杂的是,资产评估涉及公司内部财务数据、客户信息或市场交易数据等敏感信息,而这些信息的获取受到法律法规严格限制,使得数据共享面临阻碍,企业在构建资产评估模型时往往难以获得足够全面的数据,从而限制了模型的训练效果和应用范围,数据隐私问题还带来了合规性风险,增加了模型开发和应用的复杂性,这些因素相互作用,使得资产评估模型构建在数据质量和获取方面面临极大挑战,直接影响了模型的精确度和稳定性。

2.2 数据特征复杂性

数据特征的复杂性在基于机器学习的资产评估模型构建中形成了一个主要挑战,资产评估过程中涉及市场环境、行业趋势、公司财务数据等多方面的特征,特征种类和维度庞杂,如市场环境特征包含宏观经济指标、利率波动、市场波动等多个因素,而行业趋势则体现在竞争格局、技术进步、政策变化等方面,公司内部财务数据还包含收入、成本、资产负债表等多项指标,且这些特征之间往往存在非线性的复杂关系,如市场环境变化会对不同公司财务表现产生不同影响,这种影响在时间上和公司特点、市场定位上都有差异,特征之间的交互和依赖关系也加剧了特征工程的难度。资产评估数据通常来源广泛,包含公开市场数据、公司报告、行业分析和第三方研究等多种渠道,给数据整合和特征提取带来了额外困难,在这些多源数据中,特征的格式、单位和范围差异显著,必须进行标准化和清洗以保证数据的统一性和可比性。机器学习模型中特征选择和提取是关键步骤,适当的特征可以大幅提高模型预测效果,但如何从多样的多源数据中识别出最有价值的特征却是模型构建中的重大难题,不当的特征选择不仅会降低模型性能,还可能引入偏差,导致模型无法准确反映资产的实际价值,特征间的复杂互动和非线性关系使得识别重要特征和捕捉其特性变得尤为困难,需要具备深厚的领域知识和丰富的经验才能应对这些挑战。

2.3 模型选择和解释性

资产评估模型在构建过程中,机器学习模型的选择及其解释性成为主要难题之一,尤其是当资产评估涉及多种算法时,包括线性回归、决策树、随机森林和神经网络等,每种算法在数据处理上的优势和不足各不相同,往往需要结合丰富的领域知识和大量实验才能确定适合的模型,算法在预测精度、训练速度和可扩展性上表现不一,因此在应用中效果会产生较大差异,导致模型选择的过程复杂而具有挑战性,深度学习等复杂模型在处理大规模数据和捕捉非线性关系时虽能带来高精度的预测,由于其“黑箱”特性,解释性不足的问题较为明显,这对资产评估至关重要,因为评估结果不仅关系到金融决策的可靠性,还涉及投资者的信任与合规要求,金融行业需要保证决策过程的高度透明,以帮助相关方理解模型的逻辑与结果,因此解释性不足的深度学习模型在实际应用中可能不易获得广泛认可,成为其进一步推广的障碍。

3 加强机器学习技术应用于资产评估模型构建的有效策略

3.1 数据质量提升与处理

数据质量的提升和处理在资产评估模型的构建中起到关键作用,确保用于训练模型的数据既准确又全面,并且没有偏差非常重要,数据的准确性要求所收集的信息真实可靠,能够反映资产的实际情况,全面性则指数据需要覆盖多种因素,以便模型可以从多个角度进行分析,同时数据的无偏性意味着在采集过程中要避免选择性偏差,以保证模型训练的公正性,在具体的数据处理中,清洗步骤格外重要,对于缺少数据的情况,需采用插值或其他补充方法来填补缺失值,以确保数据集的完整性,而存在噪声的数据则需通过筛选和过滤方式进行清除,特别是那些可能对评估结果产生显著影响的异常值,此过程不仅有助于提高模型的预测准确性,还能降低后续分析中的误差,特征选择和特征工程则是进一步提升模型性能的关键,通过合理的特征选择可以找出对评估结果影响最大的变量。从而简化模型结构并提高解释性,特征工程则通过构建新的特征来更好地捕捉数据中的潜在模式。例如,可以通过组合或转换现有变量来生成新的特征,帮助模型深入理解资产的核心驱动因素。

3.2 模型选择与集成学习

资产评估模型的构建过程中,模型的选择和集成学习对提升预测效果和保持结果稳定起到关键作用,不同的数据特点和评估需求下,通常会使用多种机器学习算法,线性回归适合处理线性关系的数据,随机森林在复杂非线性关系下表现更好,支持向量机则更适合高维特征空间问题,在实际应用中,通过交叉验证对这些模型进行测试,观察它们在训练集和验证集上的表现,从而找出最合适的算法,并为模型选择提供数据依据,此外,集成学习技术通过结合多个模型的优势进一步提高评估的精确度和稳定性,如 Bagging 方法通过对训练数据多次采样以构建多个模型,将预测结果平均或投票处理,减少方差并增强稳定性。而 Boosting 方法则通过调整样本权重,让模型逐步关注预测误差较大的样本,从而有效提升整体预测精度。

3.3 特征选择与解释性增强

资产评估模型构建中,特征选择和解释性提升是决定模型性能和用户信任度的重要步骤,模型开发者通过从众多特征中找出对输出结果最具影响的变量,提高模型的准确性和解释性,借助特征重要性评分,尤其是采用随机森林等集成学习算法的评分方式,模型能够得到直观且量化的特征贡献度评价,使评估人员清楚地识别哪些因素在资产价值中起

主导作用,进而优化模型架构并指导特征工程实施,模型透明度和解释性提升过程中,选用易于理解的模型同样重要,简单线性回归和决策树相比复杂深度学习模型更具透明度,评估人员能够清楚掌握模型的运作逻辑,用户通过观察模型输出,可以更直观地理解各特征对最终评估结果的影响,这样既提升了决策的透明度,也增强了客户对结果的信任。此外,通过结合现代解释性技术如 LIME 和 SHAP,可以进一步揭示模型内部的决策逻辑,这些方法提供单次预测的特征贡献度分数,使用户清楚特征对特定预测的具体影响,提升模型的操作便捷性和用户体验。

4 结语

机器学习技术在资产评估中的应用具有广阔的前景,但也面临数据质量、模型选择和解释性等诸多挑战。为有效利用机器学习提高资产评估的科学性,需要不断优化算法、改进数据管理,以及结合专业知识进行模型验证。通过加强技术与行业经验的结合,资产评估的智能化水平将得到显著提升,为金融市场的稳定和资源配置的优化提供有力支持。希望未来的研究能继续深化这一方向,开发出更加完善的资产评估解决方案。

参考文献:

- [1] 李云,童良君,王文博.一种基于机器学习的不良资产评估方法及系统:202410841842[P]. [2024-11-03].
- [2] 周骥.在线学习及其在智能交通与金融工程中的应用[D].上海:复旦大学,2024.
- [3] 李宇雄.基于在线集成学习的入侵检测方法研究[D].西安:西安电子科技大学,2020.
- [4] 郭瀚文.基于机器学习模型的公司债违约预测研究[D].杭州:浙江财经大学,2022.
- [5] 王晨巍,黎歆雨,高大伟,等.基于PSO-LightGBM的网络资产脆弱性评估模型[J].信息对抗技术,2023,2(2):54-65.
- [6] 钱跃磊,李京忠.国家公园生态资源资产定价研究——基于机器学习评估偏远5A级景区宗地地价[J].价格理论与实践,2023(1):131-134+203.
- [7] 任建宇.基于集成机器学习的数据资产定价模型及系统设计[J].中国管理信息化,2022(14):80-82.

作者简介:韩萌(1993-),女,中国陕西咸阳人,硕士,讲师,从事资产评估研究。

基金项目:论文系陕西服装工程学院校级线下特色课程《资产评估学》项目研究成果(项目编号:2024TSKC070)。