徐睿, 刘旭东, 姚旺君,等. 机器学习驱动的卫星过境预测技术[J]. 智能计算机与应用,2026,16(5):190-196. DOI:10.20169/j. issn. 2095-2163. 24072402

机器学习驱动的卫星过境预测技术

徐 睿, 刘旭东, 姚旺君, 张 楠, 赵 静
(华北计算机系统工程研究所, 北京 100083)

摘 要: 在当前网络安全和空间技术快速发展的背景下,卫星过境预测对于提高通信效率、支持环境监测和加强国防安全具有重要意义。本研究旨在通过机器学习方法进行卫星过境预测,通过引入时间特征的三角变换进行了数据增强与特征选择,并使用多种常见的机器学习模型进行对比训练与预测。实验结果表明,随机森林模型在准确度和泛化能力上表现最优。研究的主要贡献在于提出了一种结合机器学习和时间特征的卫星过境预测方法,不仅为卫星通信网络的监测和管理提供了强有力的支持,也为网络安全领域提供了新的技术手段。

关键词: 卫星过境预测; 数据增强; 特征选择; 机器学习; 随机森林
中图分类号: TP181 **文献标志码**: A **文章编号**: 2095-2163(2026)05-0190-07

A machine learning-driven satellite transit prediction technology

XU Rui, LIU Xudong, YAO Wangjun, ZHANG Nan, ZHAO Jing
(National Computer System Engineering Research Institute of China, Beijing 100083, China)

Abstract: In the context of the rapid development of cybersecurity and space technology, satellite transit prediction plays a significant role in enhancing communication efficiency, supporting environmental monitoring, and strengthening national defense security. This study aims to predict satellite transits using machine learning methods, introducing a trigonometric transformation of time features for data augmentation and feature selection, and employing various common machine learning models for comparative training and prediction. The experimental results show that the random forest model performs best in terms of accuracy and generalization ability. The main contribution of this research is the proposal of a satellite transit prediction method that combines machine learning and time features, providing strong support for the monitoring and management of satellite communication networks and offering new technical means for the field of cybersecurity.

Key words: satellite transit prediction; data augmentation; feature selection; machine learning; random forest

0 引 言

随着空间技术的飞速发展,卫星已成为现代通信、导航、环境监控及国防安全的核心组成部分[1]。特别是在网络安全领域,及时准确地获取和利用卫星数据对于监测网络威胁、保卫关键基础设施以及支持全球范围内的数据传输至关重要[2]。有效利用这些卫星资源的关键在于能够准确预测其过境时间和覆盖范围。精确的卫星过境预测对地面观测任务规划、数据高效收集和资源优化配置至关重要,同时对于实时网络安全监控与响应机制的建立和维护具有决定性影响。

面对这一需求,传统的预测方法通常依赖于轨道力学复杂计算及历史数据分析,这些方法尽管在某些情况下有效,但往往忽略了时间特征与地面站位置间复杂的相互作用及其对卫星可见性和过境时间的影响[3]。在网络安全应用中,这一缺陷尤为突出,因为网络威胁的实时监测和快速反应需求迫切需要更高效、更精确的预测技术以保障数据的实时性和准确性[4]。此外,在处理大规模数据集时,这些传统方法在效率和准确度上面临挑战,特别是在迫切需要实时或近实时预测的场景中。

鉴于此,本文通过机器学习模型,在指定地面站位置与时间段的情况下,进行卫星过境与否的预测,以满足网络安全领域的特定需求。本文对比分析了 5 种不同机器学习模型,包括随机森林、支持向量机

作者简介: 徐 睿(1990—),男,博士,高级工程师。主要研究方向:信息内容安全,自然语言处理。
通信作者: 赵 静(1983—),女,博士。主要研究方向:网络安全,人工智能。Email: zhaojing12321@ 126. com。

收稿日期: 2024-07-24

(Support Vector Machine，SVM)、梯度提升树(如 eXtreme Gradient Boosting，XGBoost 和 Light Gradient Boosting Machine，LightGBM)及决策树，拓展了传统基于轨道力学的预测方法。本文结合了时间特征分析与地面站位置，不仅能提升卫星资源的利用效率，还能在网络安全背景下为地面监测站提供及时、准确的数据支持，以便快速响应网络威胁并实施有效的防护措施。

本文的主要贡献如下：

(1)采用机器学习技术与时间特征分析相结合的全新视角，拓展了传统的卫星过境预测方法，为卫星过境时间预测提供了一种更全面的解决策略。

(2)创新性地应用多种机器学习模型，结合时间特征的三角变换和地面站位置，大幅提升了模型处理时间周期性和空间变量的能力，为卫星过境时间预测设定了新的标准。

(3)所开发的模型在实际应用中展现出优异性能，不仅提高了卫星资源的使用效率，也为地面观测站的任务规划和资源管理提供了有力的决策支持工具，显示了其在网络安全以及更广泛领域的应用潜力。

1　相关工作

1.1　现有的卫星过境预测方法

卫星过境预测是融合天文学、空间科学和计算机科学的综合学科领域。传统方法依赖精细的轨道力学计算，使用开普勒轨道元素和天文算法精确预测卫星相对于地面观测点的过境时间。例如，NORAD 的 SGP4 模型被广泛用于计算近地卫星的轨道[5-6]。这类基于物理的方法在精度上通常表现出色，但需依赖复杂的数学运算与深厚的天文学知识储备。

随着机器学习技术，特别是深度神经网络的发展与应用，为卫星过境预测提供了创新方案。这些模型能够利用历史轨道数据，学习时间特征中的复杂依赖性和模式，预测未来的卫星轨道[7-8]，从而使用轨道参数计算过境事件[9]。机器学习的自动学习能力特别适合处理大气拖曳和太阳辐射等非开普勒效应的影响，提高卫星轨道预测的准确性，从而为卫星路径规划和空间碎片避让等应用提供强大的支持。

1.2　机器学习及其在类似问题中的应用

本文将卫星过境预测问题界定为基于时间特征的二分类问题，机器学习领域的多种模型及特征工程技术可有效地应用于该任务。这些技术能够处理时间数据的特性，通过学习历史过境数据，预测未来某时间段内的过境事件。尽管该任务并非传统的时序预测，但基于机器学习的特征工程与模型训练过程对提升分类任务的准确性至关重要。

(1)特征工程[10]在处理基于时间的分类问题时至关重要，其核心是从原始时间字符串中提取有意义的特征，如时间段的长度、特定时间段内的历史过境频率等，这些特征有助于模型捕捉时间数据中的复杂依赖性与非线性关系。

(2)支持向量机(SVM)[11-12]在处理高维度、样本量小的数据集时展现出较强的性能，适合于从时间特征中提取的高维数据分类任务。通过优化决策边界，SVM 可以有效地区分不同类别的数据。

(3)梯度提升树(如 XGBoost[13-14]和 LightGBM[15])适用于处理具有复杂模式的分类问题，通过逐步增强模型的预测能力来最小化误差，可从时间特征中学习深层的数据关系。

(4)决策树[16]具有易于理解和解释的优势，为处理基于时间的分类问题提供了直观的方法。在集成学习框架下，决策树也能显著提升分类任务的性能。

(5)随机森林[17-18]通过集成多个决策树来处理大量的特征数据，可有效地管理特征的多样性并提供准确的分类结果。

1.3　本研究与现有工作的不同之处

本文在卫星过境预测的实践中引入了 3 项创新，区别于现有的研究工作：首先，本文采用了多种机器学习模型，并通过特定的数据预处理和特征工程方法处理时间数据；其次，不同于传统依靠轨道力学计算的预测方法，本文的模型能够直接从历史的过境时间数据中学习，避免了对复杂物理模型与专业天文学知识的依赖；此外，引入的时间特征三角变换和数据增强手段有效提升了模型处理时间数据周期性和解决数据不平衡问题的能力。

进一步地，本文采用了系统的参数调优和网格搜索策略，精细地调整机器学习模型的配置，以满足卫星过境预测的具体需求。本文的这种方法不仅显著提升了预测准确率，还增强了模型在面对不同卫星和地面站观测任务时的泛化能力。通过这些差异，本文为卫星过境预测领域提供了新的视角和工具，展示了机器学习在解决空间科学问题中的潜力。

2　方法

2.1　数据预处理和特征工程

原始数据集源自于模拟的卫星过境地面站点记录，包含大约 30 000 条记录，每条记录的格式为[(开始时间，结束时间，1)]，其中，时间格式为

“yyyy-mm-dd HH:mm:ss”,1 代表过境。为了增强模型预测的准确性与效率,本文对原始数据集进行了数据预处理、特征工程,主要涉及数据增强、提取与转换时间特征,以及对特征进行选择。通过这些步骤,本文旨在精确捕捉时间特征的动态变化并整合地面站位置,以提升模型在预测卫星过境时间方面的表现。

2.1.1 数据增强

本文的原始数据集中只包含了卫星过境的时间段(即标签为 1 的数据),导致了标签分布的严重不平衡。为了解决这一问题,并增强模型训练过程中的数据多样性,本文采用了一种数据增强方法来生成标签为 0 的数据,即卫星未过境的时间段。

数据增强通过以下步骤实现:

(1)计算总时间范围:计算开始时间和结束时间之间的总秒数,作为潜在的数据增强范围。

(2)段落划分:基于总时间范围,计算可以划分为 k 个最大持续时间为 n 秒($n = 300$)的小时间段。如果 k 过大,则除以系数 m($m = 60$)将其减少。

(3)随机生成时间段:通过随机选择开始时间点,并在此基础上加上一个随机的时间偏移(不超过 n),生成新的时间段。本过程会根据上一步骤确定的子段数量重复进行,从而生成多个未过境的时间段样本。

这种数据增强方法对于平衡数据集和提高模型的泛化能力至关重要。通过生成未过境的样本,本文不仅扩充了训练数据集,还提高了模型在处理未见过境情况时的鲁棒性。此外,这种方法通过引入随机性,帮助模型学习到更加泛化的特征,而不是仅仅依赖于特定的过境时间段特征。

2.1.2 时间特征的提取和三角变换

时间特征对于卫星过境预测至关重要,因为卫星的可见性受到地球自转和卫星轨道周期性变化的影响。首先,从原始数据中提取了以下时间特征:开始时间和结束时间的时间戳,以及这些时间戳对应的小时、日、月和年。

接着,为了捕获时间特征的周期性,对部分时间特征进行了三角变换。具体来说,本文采用了正弦和余弦函数来转换小时和星期的特征,以此来表达一天中各个时段和一周中各个日子的周期性。通过这种方法,模型可以更好地理解时间的循环性质,比如午夜与中午的关联性,以及一周的开始与结束的相似性。这种时间特征的三角变换有助于模型在预测时考虑到时间的自然周期性,从而提高预测的准确性和效率。最终,形成的特征向量见表 1。

表 1 特征及取值范围

Table 1 Features and value ranges

序号	特征	取值范围
1	卫星 ID	见 NORAD 编号
2	地点 ID	[00000~99999]
3	开始时间戳	[1672502400~1704038398]
4	结束时间戳	[1672502401~1704038399]
5	开始小时	[0~23]
6	开始日	[1~31]
7	开始月	[1~12]
8	开始年	[2023]
9	结束小时	[0~23]
10	结束日	[1~31]
11	结束月	[1~12]
12	结束年	[2023]
13	开始小时的正弦值	[-1~1]
14	开始小时的余弦值	[-1~1]
15	开始日的星期几的正弦值	[-1~1]
16	开始日的星期几的余弦值	[-1~1]
17	结束小时的正弦值	[-1~1]
18	结束小时的余弦值	[-1~1]
19	结束日的星期几的正弦值	[-1~1]
20	结束日的星期几的余弦值	[-1~1]

2.1.3 特征选择

特征选择旨在识别出对模型预测性能最有贡献的特征,不仅可以提高模型的训练效率,还能提升模型的泛化能力。为了实现这一目标,本文采用了 3 种不同的特征选择方法:皮尔森相关系数(Pearson Correlation)、最大信息系数(MIC)和随机森林重要性评估。

(1)皮尔森相关系数(Pearson Correlation)

皮尔森相关系数是衡量两个变量线性相关程度的指标。在特征选择过程中,计算了每个特征与目标变量之间的皮尔森相关系数,以识别与卫星过境事件有强线性相关的特征,相关系数热图如图 1 所示。特征与目标变量的相关系数越接近于 1 或-1,说明其之间的线性关系越强,该特征对于模型来说越重要。本文选取相关性系数大于 0.8 的 9 个特征作为初步选取结果,分别为:开始时间戳、结束时间戳、开始小时、开始日、结束月、开始小时的正弦值、开始日的星期几的正弦值、结束小时的正弦值、结束日的星期几的正弦值。

(2)最大信息系数(MIC)

最大信息系数是一种捕捉变量之间复杂关系

（包括非线性关系）的方法。与皮尔森相关系数不同，MIC 能够识别出变量之间的任何类型的关系。通过计算特征与目标变量之间的 MIC 值，可以发现哪些特征与卫星过境事件存在强相关性，即使这种相关性是非线性的，相关系数热图如图 2 所示。本文选取 MIC 系数大于 0.8 的 11 个特征作为初步选取结果，分别为：开始时间戳、结束时间戳、开始小时、开始月、结束小时、结束日、开始小时的正弦值、开始小时的余弦值、结束小时的余弦值、结束日的星期几的正弦值、结束日的星期几的余弦值。

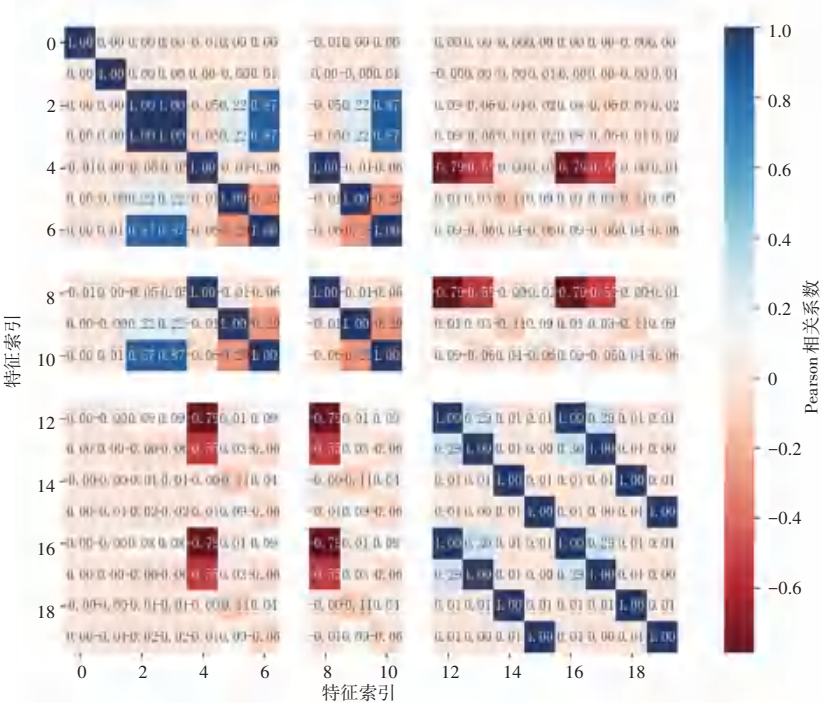


图 1　Pearson 相关系数热图

Fig. 1　Pearson correlation coefficient heatmap

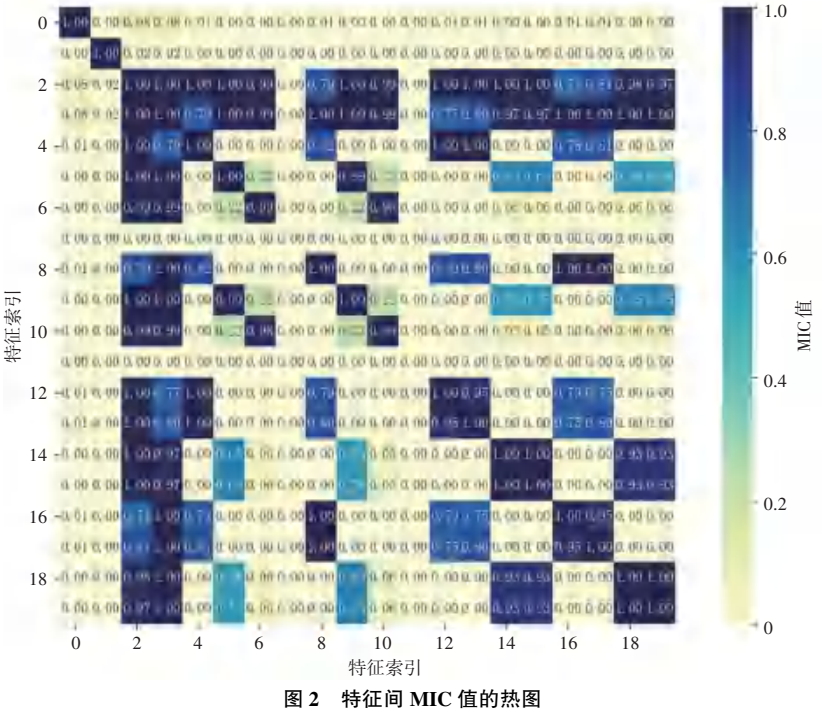


图 2　特征间 MIC 值的热图

Fig. 2　Heatmap of MIC values between features

(3)随机森林重要性评估

随机森林是一种集成学习方法,通过构建多棵决策树来进行预测。随机森林能够提供关于特征重要性的直观评估,这是通过观察每个特征在构建决策树时对模型性能的贡献程度来实现的。特征重要性越高,表明该特征在预测卫星过境事件中越重要。最终,随机森林排序出来的特征如图 3 所示。

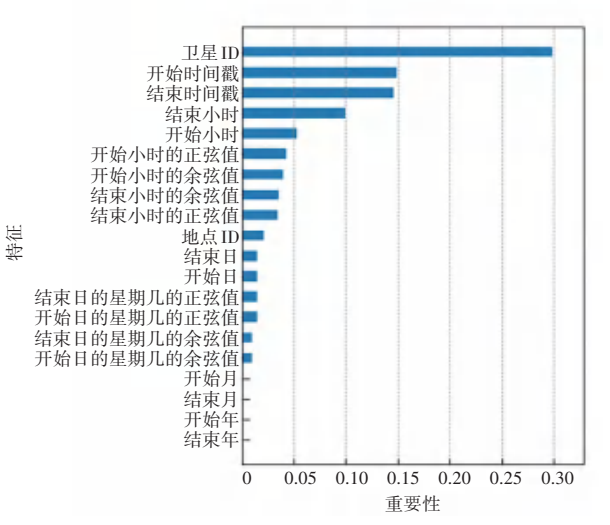


图 3 随机森林特征重要性

Fig. 3 Random Forest feature importance

本文根据上述结果做出如下特征选择。首先选取皮尔逊相关系数和最大信息系数方法的特征并集,然后与随机森林重要性评分方法的结果相交,最后加上卫星 ID 与地点 ID 得到最终的特征集合,即[卫星 ID、地点 ID、开始时间戳、结束时间戳、开始小时、开始日、开始月、结束小时、结束日、结束月、开始小时的正弦值、开始小时的余弦值、开始日的星期几的正弦值、开始日的星期几的余弦值、结束小时的正弦值、结束小时的余弦值、结束日的星期几的正弦值、结束日的星期几的余弦值]。这个特征集合综合考虑了不同特征选择方法的优点,既包含了统计相关性和互信息高的特征,也考虑了在随机森林模型中表现为重要的特征,为模型训练提供了坚实的基础。

2.2 机器学习模型的构建

在构建机器学习模型的过程中,参数选择对模型的性能产生显著影响。为了找到最佳模型配置,本文采用了网格搜索方法(GridSearchCV)来系统地遍历多种算法的参数组合。这种方法不仅适用于随机森林,也同样适用于其他机器学习算法,如支持向量机(SVM)、梯度提升树(如 XGBoost 和 LightGBM)以及决策树等,可以对每种算法进行彻底的参数优化。

以随机森林为例,本文定义了一个参数网格,包括:

(1) *n*_estimators(森林中树木的数量):从 100 到 1 000 变化,以增加模型的稳定性和性能。

(2)max_depth(树的最大深度):限制树的复杂度,防止过拟合。

(3)min_samples_split(分裂内部节点所需的最小样本数):从 2 到 10 变化,减少过拟合风险。

(4)min_samples_leaf(叶节点所需的最小样本数):从 1 到 10 变化,防止树过度生长。

(5)max_features(寻找最佳分割时考虑的特征数量):可以是“None”,“sqrt”,“log2”,或具体数字。

通过应用 GridSearchCV 并结合 5 折交叉验证,自动遍历了机器学习模型的所有参数组合,评估了每种组合的性能,最后选取最优的参数组合。

鉴于数据不平衡可能对模型性能造成显著影响,本文还采用了调整数据平衡的策略,如过采样少数类或欠采样多数类。这种策略有助于改善模型对不平衡数据的处理能力,确保预测精度和泛化能力的提升。

3 实验设计和结果

3.1 对比模型介绍

本文在对比分析中采用了多种机器学习模型,旨在评估和比较这些模型在数据集上的性能。以下是选用模型的简要介绍:

(1)支持向量机(SVM):该模型是一种广泛应用于分类和回归问题的监督学习模型。其通过在特征空间中找到最佳的线性分割超平面,以最大化正负样本之间的边缘,实现数据类别的分隔。SVM 在处理高维数据和非线性问题方面表现出色,尤其是在样本数量较少的情况下。

(2)LightGBM(LGB)[19]:该模型是一个基于梯度提升框架的高效、分布式、快速且高性能的梯度提升决策树算法。其使用基于直方图的算法,优化了大规模数据处理的速度和效率,并支持高效的并行训练。LightGBM 在处理大规模数据时,相比其他梯度提升树算法,显示出更高的效率和更快的训练速度。

(3)XGBoost[20]:该模型是一种基于梯度提升的决策树算法,其被设计用于提高速度和模型性能。XGBoost 通过优化计算资源的使用和增加正则化项来防止过拟合,在多种机器学习竞赛和任务中取得了显著的成绩。XGBoost 的核心在于能够自动利用计算机的多核处理能力,加速模型的训练过程。

(4)决策树(Decision Tree,DT):该模型是一种

基本的分类与回归方法。其通过学习简单的决策规则从数据特征中递归地构造分支，形成树状结构以进行预测。决策树易于理解和实现，能够处理数值和类别数据，并且通过树的深度可以控制模型复杂度，防止过拟合。

(5)随机森林(Random Forest，RF)：该模型是一种集成学习方法，其通过构建多个决策树并将预测结果进行汇总来提高预测的准确性。随机森林对于过拟合有很好的抵抗力，且易于处理高维数据。在多种数据集上的表现均衡，被广泛应用于分类和回归问题。

3.2　模型评价指标

在本节中，本文简要介绍用于评估机器学习模型性能的关键评价指标。这些指标不仅能够帮助本文理解模型在特定任务上的效能，而且还能为模型的进一步优化提供方向。以下是对每个评价指标的解释：

(1)准确率(Accuracy)：准确率测量的是模型预测正确的实例占总预测实例的比例。其是最直观的性能指标，反映了模型在整体数据集上的正确预测能力。

(2)精确率(Precision)：精确率关注的是模型预测为正类(即，已过境的类别)的实例中，实际上属于正类的实例所占的比例。高精确率意味着较少的假正例(即，被错误地标记为正类的负类实例)。

(3)召回率(Recall)：召回率衡量的是所有真实正类实例中，模型正确识别为正类的实例的比例。高召回率表示模型能够捕获大多数真正的正类实例。

(4) $F1$ 分数($F1$ - Score)：$F1$ 分数是精确率和召回率的调和平均值，提供了一个同时考虑精确率和召回率的单一度量。$F1$ 分数对于不平衡数据集尤其有用，因为其不会被大量的负类实例所偏倚。

(5)ROC 曲线(Receiver Operating Characteristic Curve)与 AUC 值：ROC 曲线通过绘制模型在不同阈值下的真正率(TPR)和假正率(FPR)来评估分类器区分正负类别的能力。AUC 值为 ROC 曲线图下的面积，其值越接近 1，表明模型的分类性能越好；是一个评估模型在各种阈值设置下整体表现的有效工具。

3.3　模型结果分析

模型的输入和输出在训练和预测时保持一致。具体来说，在模型训练阶段，本文对输入数据的小时和星期进行了正余弦转换，以捕捉时间的周期性特征。在模型预测时，输入数据按照同样的处理，确保模型能够根据这些时间特征输出相应的二分类结果。

模型的实验结果见表 2。随机森林和决策树取得了较好的性能，以 0.95 的准确率领先，其余指标也表现出色，证明了其在本数据集上的强大预测能力。XGBoost 表现紧随其后，尤其在召回率上达到 0.91，显示其良好的正类识别能力。尽管 SVM 经过网格搜索优化参数后仍表现最弱，这可能表示该模型与数据集特性不匹配或存在固有的局限性。LightGBM 的结果虽优于 SVM，但也未能达到前三者的水平。总的来说，随机森林和决策树对于此数据集来说是最佳选择，而 XGBoost 也表现出了不俗的分类效果。

表 2　模型实验结果

Table 2　Model experiment results

模型	准确率	精确率	召回率	$F1$
SVM	0.66	0.65	0.73	0.62
LGB	0.80	0.74	0.85	0.75
XGBoost	0.88	0.81	0.91	0.84
随机森林	0.95	0.91	0.94	0.92
决策树	0.95	0.91	0.95	0.93

模型的 ROC 曲线如图 4 所示。ROC 曲线的 AUC 值以随机森林最高，达到了 0.98，紧随其后的是决策树、XGBoost，其 AUC 值分别为 0.95、0.94，表明这三类模型在区分正负类别方面具有极高的准确性和可靠性。相比之下，LightGBM 的 AUC 值为 0.91，虽然较高，但相较于领先的模型略显不足。SVM 的 AUC 值仅为 0.50，显示出其在区分正负类别方面的能力极为有限，这与其在准确率、精确率、召回率和 $F1$ 分数上的表现相一致。综合考虑各项指标，随机森林在本数据集上的表现最为出色，决策树和 XGBoost 亦表现优异，而 LightGBM 和 SVM 需要进一步的调整或考虑其他模型以提升效果。

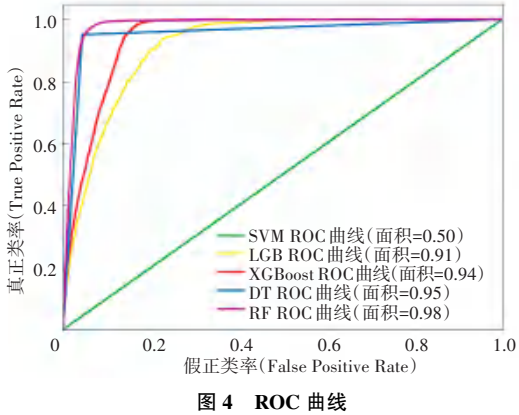


图 4　ROC 曲线

Fig. 4　ROC curve

4 结束语

本文通过对多种机器学习模型在卫星过境预测任务上的性能进行详尽的比较和分析,得出了一系列重要的结论。首先,随机森林模型因其优异的准确度和泛化能力,在所有测试的模型中表现最佳。其次,时间特征的三角变换对提升预测准确性起到了重要作用,这表明在卫星过境预测中考虑时间周期性是十分必要的。此外,数据增强和特征选择过程对于处理原始数据集中的不平衡与噪声干扰,提高模型的预测性能也显示出了其重要性。

实验结果还揭示了不同机器学习模型对于卫星过境预测问题的适用性差异,这提示未来研究可以进一步探索模型结合或融合的策略,以利用各模型的优势。综上所述,本文展示了机器学习技术在提高卫星过境预测精度方面的潜力,并为网络安全领域的应用提供了有力的技术支持。未来工作可以考虑更多的特征工程技术和深度学习模型,比如将地面站位置转换为特征、使用 CNN 和 LSTM 等,进一步探索卫星过境预测的精确性提升空间。

参考文献

[1] RAHMAT-SAMII Y, DENSMORE A C. Technology trends and challenges of antennas for satellite communication systems[J]. IEEE Transactions on Antennas and Propagation, 2014, 63(4): 1191-1204.

[2] AHMAD I, SUOMALAINEN J, PORAMBAGE P, et al. Security of satellite-terrestrial communications: Challenges and potential solutions[J]. IEEE Access, 2022, 10: 96038-96052.

[3] LEVIT C, MARSHALL W. Improved orbit predictions using two-line elements[J]. Advances in Space Research, 2011, 47(7): 1107-1115.

[4] LI K, ZHOU H, TU Z, et al. Distributed network intrusion detection system in satellite-terrestrial integrated networks using federated learning[J]. IEEE Access, 2020, 8: 214852-214865.

[5] 王密, 陈俊博, 皮英冬,等. 一种面向卫星在轨自主任务规划的快速精准轨道预报方法[J]. 武汉大学学报(信息科学版), 2024, 49(6): 879-887.

[6] 张奔, 陈祥, 孙华. 低轨卫星轨迹预测和可视化研究[J]. 信息化研究, 2019, 45(3): 40-45.

[7] 张心宇, 刘源, 宋佳凝. 基于 LSTM 神经网络的短期轨道预报[J]. 系统工程与电子技术, 2022, 44(3): 939-947.

[8] 朱俊鹏, 赵洪利, 杜鑫,等. 长短时记忆神经网络在卫星轨道预报中的研究[J]. 兵器装备工程学报, 2017, 38(10): 127-132.

[9] 吕亮, 赵英豪, 徐青,等. 一种遥感卫星连续过境区域快速判定方法[J]. 测绘科学技术学报, 2018, 35(2): 165-169.

[10] 胡静静, 樊军. 多特征加权集成模型在用户购买预测中的研究[J]. 机械设计与制造, 2024,48(4): 373-378.

[11] 李贝贝, 孙深圳, 刘秀梅,等. 基于改进 SVM 算法的调节阀空化状态识别[J]. 机械工程学报, 2023, 59(24): 312-322.

[12] 闫芳序, 王剑辉. 基于 SVM 和 Word2vec 的微博评论情感识别模型[J]. 现代计算机, 2024, 30(10): 60-64.

[13] 赵军, 李红玮, 刘小杰,等. 基于 Xgboost 的高炉透气性指数预测模型[J]. 中国冶金, 2021, 31(3): 22-29.

[14] 方茜. 基于 XGBoost 的网络舆情文章预警模型建立[J]. 智能计算机与应用, 2023, 13(6): 72-77.

[15] 马天鹤, 侯本伟, 缪惠全,等. 基于 LightGBM 算法的供水管网抗震性能评估[J/OL]. 工程力学(2024-02-19)[2026-02-10]. https://engineeringmechanics.cn/cn/article/id/1ad85aaf-ea96-4889-bb8b-4daf4d2fbd18.

[16] 丁明彬. 基于小波变换和决策树的电机滚动轴承故障诊断[J]. 内燃机与配件, 2023,44(23): 54-57.

[17] 卢飞, 宋佳佳. 随机森林算法在擦机尾事件预测中的应用研究[J]. 安全与环境学报, 2024, 24(8): 3028-3035. https://doi.org/10.13637/j.issn.1009-6094.2023.2121.

[18] 徐英卓, 王若禹, 王六鹏. 基于遗传算法优化的随机森林钻井机械钻速预测模型研究[J]. 智能计算机与应用, 2023, 13(3): 210-214.

[19] KE G, MENG Q, FINLEY T, et al. Lightgbm: A highly efficient gradient boosting decision tree[C]//Proceedings of the Advances in Neural Information Processing Systems. NeurIPS, 2017: 3149-3157.

[20] CHEN T, GUESTRIN C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd Acm SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 785-794.