



基于数据挖掘的通信网络故障分类研究

朱圳, 刘立芳, 齐小刚

引用本文:

朱圳,刘立芳,齐小刚. 基于数据挖掘的通信网络故障分类研究[J]. 智能系统学报, 2022, 17(6): 1228–1234.

ZHU Zhen,LIU Lifang,QI Xiaogang. Research on communication network fault classification based on data mining[J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(6): 1228–1234.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202111037>

您可能感兴趣的其他文章

基于双向LSTM卷积网络与注意力机制的自动睡眠分期模型

Automatic sleep staging model based on the bi-directional LSTM convolutional network and attention mechanism

智能系统学报. 2022, 17(3): 523–530 <https://dx.doi.org/10.11992/tis.202103013>

利用置信规则库构建WSN节点故障检测模型

Constructing a WSN node fault detection model using the belief rule base

智能系统学报. 2021, 16(3): 511–517 <https://dx.doi.org/10.11992/tis.202009006>

移动通信网络的中性集故障诊断方法研究

Research on neutral set fault diagnosis method for mobile communication networks

智能系统学报. 2020, 15(5): 864–869 <https://dx.doi.org/10.11992/tis.201906031>

网络拓扑特征的不平衡数据分类

Imbalanced data classification of network topology characteristics

智能系统学报. 2019, 14(5): 889–896 <https://dx.doi.org/10.11992/tis.201812014>

基于粗糙集相对分类信息熵和粒子群优化的特征选择方法

A feature selection approach based on rough set relative classification information entropy and particle swarm optimization

智能系统学报. 2017, 12(3): 397–404 <https://dx.doi.org/10.11992/tis.201705004>

DOI: 10.11992/tis.202111037

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20221013.1334.004.html>

基于数据挖掘的通信网络故障分类研究

朱圳¹, 刘立芳¹, 齐小刚²

(1. 西安电子科技大学 计算机科学与技术学院, 陕西 西安 710071; 2. 西安电子科技大学 数学与统计学院, 陕西 西安 710071)

摘 要: 以往针对通信网络故障分类的算法没有考虑告警和故障数据中的潜在特征, 导致故障分类准确率低, 因此提出一种基于数据挖掘的通信网络故障分类算法。首先, 根据对数据背景和数据特点的理解, 使用特征构造挖掘数据中潜在的特征, 将挖掘到的特征加入原数据中。然后, 使用 LightGBM 算法的特征重要性评估函数对新数据集中的所有特征进行重要性评估, 根据重要性值删除不重要特征。最后, 使用集成学习模型对特征筛选后的数据集进行故障分类研究。实验结果表明, 基于数据挖掘的通信网络故障分类算法的准确率有更好的效果。

关键词: 数据挖掘; 通信网络; 故障分类; 集成学习; 特征工程; 告警数据; 网络故障; 网络告警

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2022)06-1228-07

中文引用格式: 朱圳, 刘立芳, 齐小刚. 基于数据挖掘的通信网络故障分类研究 [J]. 智能系统学报, 2022, 17(6): 1228-1234.

英文引用格式: ZHU Zhen, LIU Lifang, QI Xiaogang. Research on communication network fault classification based on data mining[J]. CAAI transactions on intelligent systems, 2022, 17(6): 1228-1234.

Research on communication network fault classification based on data mining

ZHU Zhen¹, LIU Lifang¹, QI Xiaogang²

(1. School of Computer Science and Technology, Xidian University, Xi'an 710071, China; 2. School of Mathematics and Statistics, Xidian University, Xi'an 710071, China)

Abstract: The communication network fault classification algorithms previously used did not consider potential features in alarm and fault data, resulting in low classification accuracy. This study proposes a data mining-based communication network fault classification algorithm to address this issue. First, the feature structure is applied to the mining of the potential features in the data based on the understanding of the data background and data characteristics, and then the mined features are added to the original data. Furthermore, the LightGBM algorithm's feature importance evaluation function is used to evaluate the importance of all features in the new dataset and to delete unimportant features based on the importance value. Finally, ensemble learning algorithms are used to perform fault classification studies on feature-screened datasets. The experimental results show that the accuracy of the communication network fault classification algorithm based on data mining has a better effect.

Keywords: data mining; communication network; fault classification; integrated learning; characteristic engineering; alarm data; network failure; network alarm

随社会的快速发展, 对于网络的需求也越来越大。智能手机、智能设备、智能家居等的出现, 增大了网络的使用。传统通信网络面临着前所未

有的增长, 对网络的需求和使用也在增大, 导致网络负担变大。因此通信网络经常发生故障, 故障以告警的形式进行上报, 一旦某处发生网络故障, 网络中就会产生大量告警信息, 如何快速定位网络故障类型是一个难题, 而传统研究基本都在关注告警之间的关联规则。在文献 [1] 中将群

收稿日期: 2021-11-19. 网络出版日期: 2022-10-14.

基金项目: 国家自然科学基金项目 (61877067).

通信作者: 刘立芳. E-mail: lfliu@mail.xidian.edu.cn.

智能算法用于关联规则的挖掘,并应用于通信领域。文献[2]同样利用群智能算法中的蚁群算法进行告警的关联分析。文献[3-5]都是基于频繁模式树的关联规则方法,该方法提高了算法的运行效率。除关联规则模式挖掘之外,序列模式也频繁地应用在通信告警领域,序列模式挖掘考虑时间上的顺序,从而效果更佳。文献[6]使用的是序列模式挖掘。文献[7-8]都是基于序列模式挖掘的实际应用。除研究通信网络的告警关联和序列模式,还有一些国内外的研究学者研究网络的故障定位,也取得了一些不错的成果。2002年Steinder等^[9]提出一种基于贝叶斯网络的故障定位技术,贝叶斯网络根据网络拓扑和通信协议构建,并使用Pearl's iterative算法进行概率推理,但该方法只能用于单连通网络。许多学者引入贝叶斯网络为网络故障事件的关系建立模型^[10]。王开选等^[11]指出了故障传播模型下的故障定位问题是NP困难(non-deterministic polynomial)问题,并提出一种启发式的最小损失故障定位算法。同时故障诊断系统也被开发和使用,华为诺亚方舟实验室开发了诊断系统,通过对历史数据的分析和构建知识图谱,并根据知识图谱进行推理,可以以问答的形式辅助工程师找到故障根因^[12]。王迎春等^[13]使用规则进行故障定位,规则使用条件-结果的语句形式表示,该定位方法主要需要解决规则知识库构建的问题。初始进行故障诊断多数依赖专家,根据专家经验和网络资源等关联性建立故障推理树,并完成故障定位^[13]。赵灿明等^[14]采用二分图模型考虑了通信网络中故障位置对告警信息的影响,旨在解决大范围的故障告警下故障定位问题。

为实现通信网络故障分类,本文提出基于数据挖掘的通信网络告警分类算法。首先,针对干净的告警数据和故障数据,对其进行特征工程,将挖掘到的特征与原数据合并,得到便于模型训练的数据集;然后,基于集成学习模型对数据集进行模型训练与预测,并与文献[15]中的基于卷积神经网络故障分类进行对比实验,最终获得通信网络故障类型。

1 相关概念

1.1 数据挖掘

数据挖掘^[16]是指从大量的数据中通过算法搜索隐藏于其中信息的过程。数据挖掘分为有标签和无标签挖掘两大类。有标签数据可以用来进行分类或者预测任务,无标签数据可以用来进行

聚类或者关联分析等。

1.2 特征工程

特征工程^[17]是指将数据转换为能更好地表示潜在问题的特征的方法,从而提升机器学习的性能。特征工程就是对数据的特征或者数据进行分析,将数据转换成可以更好地表示问题的潜在特征,从而提高机器学习的性能。特征工程主要包括以下几个重要的作用。

1)转换数据格式。这也是数据预处理中一个重要的环节,但这里的转换数据不仅仅针对干净数据,也针对脏数据。有些数据以表格为主,无法直接拿来预处理,需要转换数据格式,这也属于特征工程的范畴。

2)确定特征。原始的数据中可能存在多列属性,但并非所有的属性都可以用作模型训练的特征,特征是可以标识问题的重要属性,而不能标识问题的属性称为普通属性。

3)提高学习性能。特征工程最大的作用就是获取最佳的数据,最佳的数据可以更好地标识问题,进行机器学习训练时可以得到更好的效果。

1.3 集成学习

集成学习^[18]是将若干个基学习器(分类器、回归器)组合之后产生一个新的学习器。相比单一模型,集成学习模型在准确性、稳定性、鲁棒性和泛化能力上都有很好的效果。一般来说,集成学习可以分为3类:1)减少方差(Bagging),即防止过拟合;2)减少偏差(Boosting),即提高训练样本正确率;3)提升预测结果(Stacking),即提高验证精度。

1) Bagging: 通过对样本数据集进行有放回地重复采样,生成多个采样子集,并行地训练出多个模型,测试阶段集成多个模型的泛化输出,常常采样直接平均的做法。Bagging执行流程如图1所示。

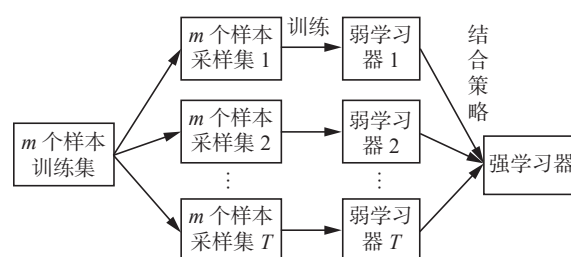


图1 Bagging执行流程

Fig. 1 Bagging execution process

2) Boosting: 其思想是采用串行训练过程来训练模型。同样是利用数据集来训练多个模型,但Boosting的最大特征是后训练的模型会考虑前训

训练模型的误差,具体做法就是对于前训练模型中出错的样本加大权重,称为赋权法。赋权法的应用使得每个样本对于训练模型的误差起到的作用是不同的,而后训练模型会采用贪心算法去不断适应训练集,力争将每个训练样本的误差都尽量降低。Boosting 执行流程如图2所示。

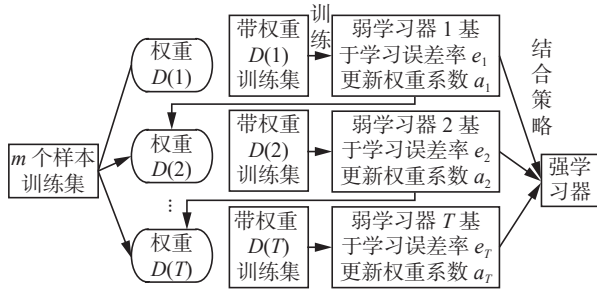


图2 Boosting 执行流程

Fig. 2 Boosting execution process

3) Stacking: 该方法是将多个不同基学习器得到的输出作为输入,训练一个新模型,得到最终结果。具体过程如下:

- ①将训练数据集随机划分为两个数据集;
- ②一个用于训练多个基学习器,一个用于测试这几个基学习器;
- ③将②得到的预测结果作为输入,训练1个更好的分类器。在第2个集合上测试这几个学习器。

2 基于数据挖掘的特征构造法

2.1 数据处理与特征构造

从数据中提取出可以用于模型训练的数据特征,比如将时间做处理,时间可以提取出年、月、日等,并且还可以根据时间来判断当前是工作日还是休息日,不同的时间点网络的负担情况是不一样的,这些因素都可能影响网络的质量情况。分析网络告警标题,由于标题是文本数据,可以对其进行文本的提取处理,可能某些告警标题就是对应着相关的故障类型,这样的告警标题更能表征故障类型。还可以根据告警或者故障发生的基站或小区名称进行分组处理,得到一些其他的特征信息^[19-24],比如某个小区出现故障或者告警的频率、故障发生的时间等。

1) 针对告警标题做 TF-IDF

告警标题:故障发生时所上报的告警名称,不同的告警标题表示不同的告警类型,如 ETH_LOS 表示端口接收不到信号、RHUB 与 pRRU 间链路异常告警和用户面故障告警表示接口异常。

TF-IDF: 一种用于信息检索与数据挖掘的常用加权技术,TF 表示某个词出现的频率,IDF 表示逆文本频率指数。IDF 的主要思想是:如果包

含某个词的文档越少,IDF 越大,则说明该词具有很好的类别区分能力。

使用 TF-IDF 对告警标题进行转换,对每个告警标题中的词统计,将文本信息换为数值信息。处理后的告警数据变成表1的形式。为了方便记录生成的词特征,用idf_i表示第i个词。

表1 TF-IDF 处理后新增的特征
Table 1 New features after TF-IDF processing

idf_1	idf_2	...	idf_92
0	0.107 860 834	...	0.384 721 336
0.048 164 473	0.237 602 251	...	0

2) 处理告警的时间特征

告警数据的时间特征主要是考虑告警的发生时间信息,将告警的发生时间进行处理,提取出告警发生时所在的月份、是在工作日还是周末等。处理后的告警时间特征如表2所示。

表2 告警数据时间处理后特征
Table 2 Characteristics of alarm data after time processing

每周平均产生告警的概率	...	告警发生的时间/h
0.042 253 521	...	9
0.302 325 581	...	14

3) 处理故障时间特征

当网络发生故障时,统计故障发生日期、故障发生的时间是否在周末、故障发生的所在的时间段、是否在工作日和故障持续时间等特征。处理后的故障时间特征如表3所示。

表3 网络故障时间处理后特征
Table 3 Characteristics after network fault time processing

故障发生的日期	...	故障发生的时间段
3	...	1
3	...	2

4) 比率特征

比率特征是将前面处理后得到的特征进行求比例,比如:求每个告警标题出现的次数,当前故障中告警标题的种类,求告警标题在每个小区的比例情况,求解每个小时发生告警的比例情况等。处理后的特征如表4所示。

表4 相关特征的比例特征
Table 4 Proportional features of related features

告警标题出现的比例	...	标题在不同小区比例
0.027 777 778	...	0.333 333 333
0.172 413 793	...	0.4

将上述这些经过特征处理后的所有特征进行合并,得到最终的训练集数据,并将得到的最新数据集用到集成学习模型。

2.2 特征选择

经过数据处理和特征构造后共得到 183 个属性,并不能将所有的属性作为特征加入到模型中进行训练,有些属性可能会影响模型的效果。本文所使用的数据量近 20 000 个样本,每个样本有 83 个特征的数据量,如果全部用于训练,将对机器要求非常高,同时算法运行时间也较慢,因此需要对属性进行筛选,使用 LightGBM 模型进行特征筛选。

LightGBM^[25] 是 2017 年由微软团队开源的集成学习模型,该模型是对梯度提升树优化的模型。该模型训练速度快、内存占用小,被广泛运用在数据科学竞赛中。该模型可以用来评估特征的重要性,对数据训练后,可以通过模型的 `feature_importance()` 函数获取特征的重要性值,该函数对训练完的各特征进行重要性排序。特征筛选流程如图 3 所示。

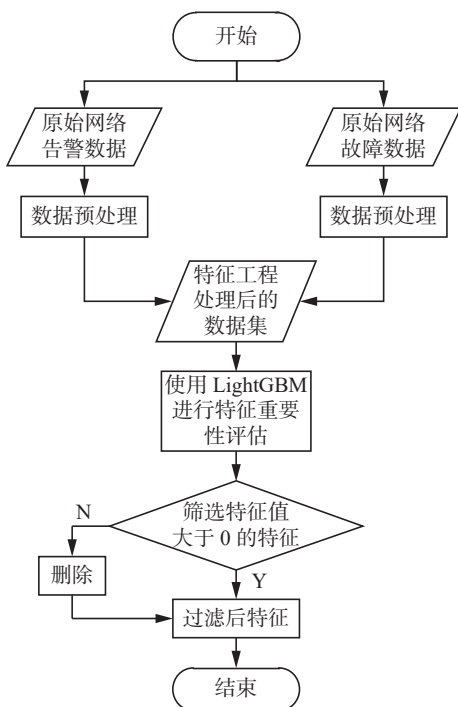


图 3 特征筛选流程图

Fig. 3 Feature screening flow chart

经 LightGBM 的重要性评估后,有 24 个特征的重要性值为 0,说明这些特征对最终分类结果没有作用,将这些特征剔除。通过特征重要性函数可以发现特征重要性值为 0 的特征多数为时间相关的特征,如告警发生的小时、告警发生是否在周末等,说明时间特征对故障分类的重要性较低。而告警标题经过 TF-IDF 处理后得到的特征,

特征的重要性值较高,说明告警标题对故障分类有着重要的作用。

3 实验与分析

为验证提出算法的性能情况,本部分通过实验进行性能分析。与文献 [15] 中的基于卷积神经网络的故障分类算法进行对比实验,通过实验分析可以看出提出的基于数据挖掘的通信网络故障分类算法有更高的分类准确率,且时间也相对比 CNN 快,因此提出的方法在故障分类的准确率上是有优势的。所有实验均在带有 8RAM 和 1T 硬盘的 Interi(R)、Core(TM)i5-4790 CPU@3.6 GHz 的计算机上进行,并使用 Python 语言和 Java 语言一起实现。

3.1 实验数据集

实验所使用的数据如表 5 所示,其主要包括电力、硬件、软件、传输和动环故障五大故障。告警序列 Alarm_i 表示第 i 个告警,其中每个告警中又包含告警发生和告警清除的时间、告警标题名称、告警发生站点等信息。

表 5 实验数据
Table 5 Experimental data

时间	故障类型	告警序列
2018/7/1 11:10:20	电力故障	Alarm ₁ , Alarm ₂
2018/7/1 11:30:33	硬件故障	Alarm ₂ , Alarm ₄
⋮	⋮	⋮
2018/7/1 16:48:38	软件故障	Alarm _j , Alarm _i
⋮	⋮	⋮

3.2 算法的评价指标

分类是机器学习中常见的任务,常见的评价指标有准确率、精确率、召回率、 F_1 -score、ROC 曲线等。混淆矩阵如表 6 所示,其中 TP (true positive) 表示真正类,即样本为正且预测也为正; FN (false negative) 表示假负类,即样本为正预测为负; FP (false positive) 表示假正类,即样本为负预测为正; TN (true negative) 表示真负类,即样本为负且预测为负。

表 6 数据检测结果
Table 6 Data test result

结果	被检测为正常	被检测为异常
正常数据	TP	FP
异常数据	FN	TN

准确率为

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

准确率是分类问题中直观的评价指标,有明显弊端,在各分类样本比重不均匀时,占比较大的分类会影响准确性的评价。

精确率为

$$P = \frac{TP}{TP + FP} \quad (2)$$

精度率是描述分类器不将负样本预测为正样本的能力。

召回率为

$$R = \frac{TP}{TP + FN} \quad (3)$$

召回率是描述分类器找出全部真正样本的能力。

F_1 -score 为

$$F_1 = \frac{PR}{P + R} \quad (4)$$

式中: P 代表精确度; R 代表召回率; F_1 -score 越大,说明模型越稳定。

3.3 算法的高效性验证

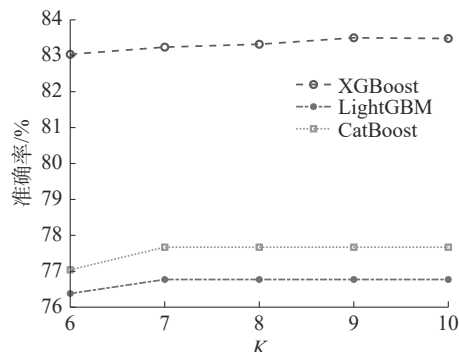
图4(a)中,对比在新数据集下3种不同集成学习模型的分类准确率差异,可以得到在不同 K 折交叉验证下,每个模型的分类准确率都在提高。当 $K \geq 7$ 时,LightGBM和CatBoost的分类准确率基本趋于不变,而XGBoost的分类准确率随 K 值变大而变大;当 $K \geq 9$ 时,XGBoost的分类准确率也趋于稳定,基本都在83.50%。从图中可明显看出在新数据集下,XGBoost的分类准确率比LightGBM和CatBoost模型的分类准确率高,而CatBoost的分类准确率又高于LightGBM。综上所述,3种集成学习模型的分类准确率大小分别是XGBoost>CatBoost>LightGBM。

考虑到数据类型存在一定的不均衡性,从模型的 F_1 -score值来对比一下3种集成学习模型的性能差异。在图4(b)中可以看出在不同 K 值下3种集成学习模型的 F_1 -score也不同,当 K 不断增大时只有XGBoost模型的 F_1 -score在不断变化,而LightGBM和CatBoost模型的 F_1 -score没有变化。对于XGBoost集成学习模型,随着 K 值的增加 F_1 -score越来越大,说明模型越来越稳定,当 $K=10$ 时, $9=F_1\text{-score}<K$ 的值,说明模型的稳定性下降了。为了说明XGBoost集成学习模型在 $K=9$ 时的效果最佳,对XGBoost模型多做几组 $K>10$ 的实验,进一步对比 K 值对该模型的影响。

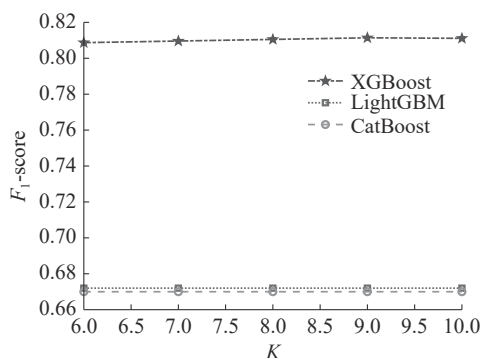
从图5可以看出,当 $6 \leq K < 9$ 时,XGBoost模型的 F_1 -score值越来越大;当 $K > 9$ 时,XGBoost模型的 F_1 -score值越来越小;当 $K=9$ 时,XGBoost模型稳定性最好。

为了证明本文提出的基于数据挖掘的方法比

文献[15]中基于卷积神经网络的网络故障分类效果好,进一步做对比实验,将新数据集在集成学习模型上的结果和文献[15]中基于CNN的结果进行对比分析。



(a) 不同 K 值下的准确率



(b) 不同 K 值下的 F_1 score

图4 不同 K 值下的结果

Fig. 4 Results under different K values

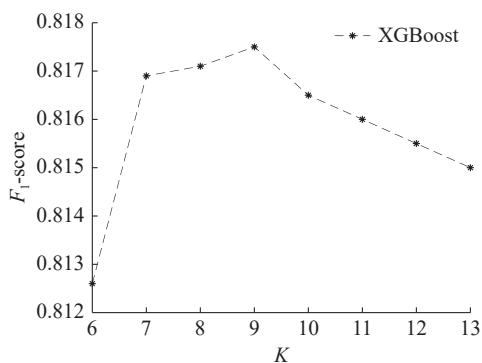


图5 不同 K 值下的 F_1 -score

Fig. 5 F_1 -score under different K values

从图6(a)可以看出,随着训练集数据量增大,基于CNN网络故障分类算法和本文提出的方法在不同集成学习模型下的分类准确率都在提高,在相同训练集数据量时,本文提出的方法在XGBoost集成学习模型下的分类准确率高于CNN算法。CNN算法的分类准确率高于LightGBM和CatBoost集成学习模型,说明基于CNN的网络故障分类算法有一定的效果。如果在数量集足够多的情况下,可能CNN算法的分类效果会更好,但

是由于数据量有限,就目前数据来看, XGBoost 集成学习模型的分类准确率更高。其主要原因是:本文提出的基于数据挖掘的网络故障分类算法考虑告警和故障之间的潜在特征,并将告警相关的特征进行了处理,挖掘到的特征可以更好地区分故障类别。而文献 [15] 中提出的基于卷积神经网络的故障分类算法,并没有考虑告警标题、时间等潜在信息,只是将告警和故障根据时间进行划分,所以信息挖掘不充分。

图 6(b) 中,对比几种模型运行时间的差异,从图可得,随数据量增大,所有模型的运行时间都增大。4 种模型的运行时间: CatBoost>CNN>XGBoost>LightBG。虽然 LightGBM 的运行时间最短,但准确率最小。而新数据集在 XGBoost 集成学习模型下的运行时间小于 CNN,因此,本文提出的方法所得到的新数据集在 XGBoost 这种集成学习模型下的效果最好。

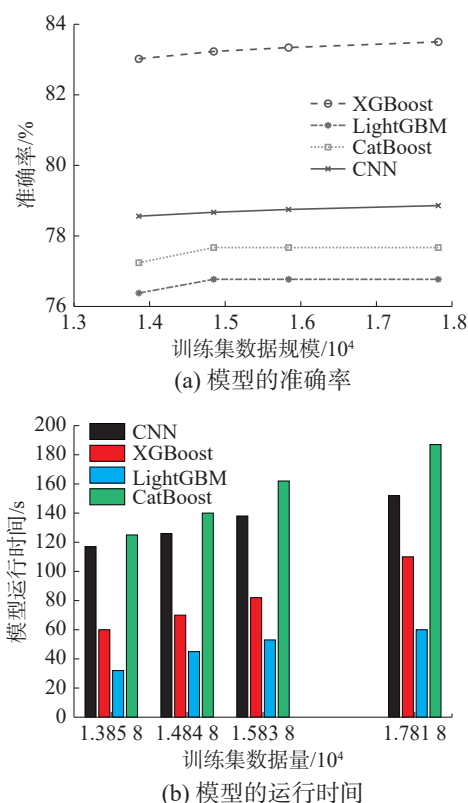


图 6 不同数据规模下的结果

Fig. 6 Results under different data scales

综上所述,本文提出的基于数据挖掘方法得到的新数据集,在 XGBoost 集成学习模型上有更好的分类准确率和更快的分类结果,可以用于通信网络故障分类。但所提出的方法也存在一定的缺点,所使用的数据量有局限,未来如果可以获取到更多的有效数据集,可以再做进一步的研究。

4 结束语

本文根据通信网络告警数据和网络故障数据,进行数据挖掘和特征构造,得到一些潜在的特征信息,将潜在特征与原数据一同进行模型的训练与预测,从结果可以看出得到的新数据集在 XGBoost 集成学习模型上的分类准确率更高。其次,从 XGBoost 和 CNN 算法的运行时间来看, XGBoost 的运行时间更短,可以在短时间内得到网络的故障类型。

参考文献:

- [1] WANG Yang, LI Guocai, XU Yakun, et al. An algorithm for mining of association rules for the information communication network alarms based on swarm intelligence[J]. Mathematical problems in engineering, 2014, 2014: 1-14.
- [2] WANG Yang, XU Yakun, LI Gang, et al. Association rules mining algorithm using sharing mechanism niche leaping ant colony based on apriori algorithm[J]. Journal of computational and theoretical nanoscience, 2015, 12(10): 3223-3231.
- [3] LI Chuan, HUANG Xiaojun. Research on FP-Growth algorithm for massive telecommunication network alarm data based on Spark[C]//2016 7th IEEE International Conference on Software Engineering and Service Science. Beijing: IEEE, 2016: 875-879.
- [4] WANG Jiantao, HE Caifeng, LIU Yijun, et al. Efficient alarm behavior analytics for telecom networks[J]. Information sciences, 2017, 402: 1-14.
- [5] 罗明, 孟传伟, 黄海量. 基于加权频繁模式树的通信网络告警规则挖掘方法 [J]. 计算机工程, 2016, 42(4): 190-196.
LUO Ming, MENG Chuanwei, HUANG Hailiang. Alarm rule mining method in telecommunication network based on weighted frequent pattern-tree[J]. Computer engineering, 2016, 42(4): 190-196.
- [6] AGRAWAL R, SRIKANT R. Mining sequential patterns[C]//Proceedings of the eleventh international conference on data engineering. Taipei: IEEE, 1995: 3-14.
- [7] WANG Jia, LI Hongguang, HUANG Jingwen, et al. Association rules mining based analysis of consequential alarm sequences in chemical processes[J]. Journal of loss prevention in the process industries, 2016, 41: 178-185.
- [8] LIM H K, KIM Y, KIM M K. Failure prediction using sequential pattern mining in the wire bonding process[J]. IEEE transactions on semiconductor manufacturing, 2017, 30(3): 285-292.
- [9] STEINDER M, SETHI A S. End-to-end service failure diagnosis using belief networks[C]// NOMS 2002. IEEE/IFIP Network Operations and Management Symposium. Management Solutions for the New Communications World' (Cat. No. 02CH37327). Florence: IEEE,

- 2002: 375–390.
- [10] CHAN A, MCNAUGHT K R. Using Bayesian networks to improve fault diagnosis during manufacturing tests of mobile telephone infrastructure[J]. *Journal of the operational research society*, 2008, 59(4): 423–430.
- [11] 王开选, 杨峥, 邱雪松. 面向影响分析的电力通信网故障定位算法[J]. *北京邮电大学学报*, 2014, 37(S1): 55–59.
- WANG Kaixuan, YANG Zheng, QIU Xuesong. Algorithm of fault locating on impact analysis in power communications network[J]. *Journal of Beijing University of Posts and Telecommunications*, 2014, 37(S1): 55–59.
- [12] SHI Jingjing, YANG Lin, ZHU Axing, et al. Machine-learning variables at different scales vs. knowledge-based variables for mapping multiple soil properties[J]. *Soil science society of America journal*, 2018, 82(3): 645–656.
- [13] 王迎春, 王远波, 王平. 基于规则的通信网故障自动诊断系统[J]. *计算机与现代化*, 2010(2): 22–25.
- WANG Yingchun, WANG Yuanbo, WANG Ping. Rule-based automatic fault diagnosis system for communication networks[J]. *Computer and modernization*, 2010(2): 22–25.
- [14] 赵灿明, 李祝红, 陶磊, 等. 基于故障传播模型与监督学习的电力通信网络故障定位[J]. *计算机应用*, 2016, 36(4): 905–908, 922.
- ZHAO Canming, LI Zhuhong, TAO Lei, et al. Fault localization for electric power communication network based on fault propagation model and supervised learning[J]. *Journal of computer applications*, 2016, 36(4): 905–908, 922.
- [15] 王菲. 基于深度学习的光网络故障定位技术研究[D]. 北京: 北京邮电大学, 2019.
- WANG Fei. Research of fault location in optical networks based on deep learning[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [16] 张曾莲. 基于非营利性、数据挖掘和科学管理的高校财务分析、评价与管理研究[M]. 北京: 首都经济贸易大学出版社, 2014.
- [17] 张银杰, 揣锦华, 翟晓惠. 基于特征工程与集成学习的恶意软件预测研究[J]. *计算机时代*, 2020(7): 7–11.
- ZHANG Yinjie, CHUAI Jinhua, ZHAI Xiaohui. Research on malware prediction based on feature engineering and ensemble learning[J]. *Computer era*, 2020(7): 7–11.
- [18] 胡毅, 瞿博阳, 梁静, 等. 进化集成学习算法综述[J]. *智能科学与技术学报*, 2021, 3(1): 18–33.
- HU Yi, QU Boyang, LIANG Jing, et al. A survey on evolutionary ensemble learning algorithm[J]. *Chinese journal of intelligent science and technology*, 2021, 3(1): 18–33.
- [19] 汪星, 黄小瑜, 刘瑄璞, 等. 面向工业大数据的多层增量特征提取方法[J]. *西安电子科技大学学报*, 2018, 45(4): 106–111.
- WANG Xing, HUANG Xiaoyu, LIU Xuanpu, et al. Multi-layer incremental feature extraction method for industrial big data[J]. *Journal of Xidian University*, 2018, 45(4): 106–111.
- [20] WANG Jing, WANG Huaichao, ZHANG Jiyang. Alarm association rules mining based on Run log for civil aviation information system[C]//2017 8th IEEE International Conference on Software Engineering and Service Science. Beijing: IEEE, 2017: 836–841.
- [21] CHEN Kuoming, CHANG T H, WANG Kaicheng, et al. Machine learning based automatic diagnosis in mobile communication networks[J]. *IEEE transactions on vehicular technology*, 2019, 68(10): 10081–10093.
- [22] JI Xilin, SHI Xiaodan, HAN Jinxi, et al. The alarm feature analysis algorithm for communication network[M]//Proceedings of the 9th International Conference on Computer Engineering and Networks. Singapore: Springer Singapore, 2020: 255–265.
- [23] WANG Danshi, LOU Liqi, ZHANG Min, et al. Dealing with alarms in optical networks using an intelligent system[J]. *IEEE access*, 2019, 7: 97760–97770.
- [24] RUIZ M, FRESI F, VELA A P, et al. Service-triggered failure identification/localization through monitoring of multiple parameters[C]//ECOC 2016: 42nd European Conference on Optical Communication. Düsseldorf: VDE, 2016: 1–3.
- [25] KE G, MENG Q, FINLEY T, et al. Lightgbm: a highly efficient gradient boosting decision tree[J]. *Advances in neural information processing systems*, 2017, 30: 3146–3154.

作者简介:



朱圳, 硕士研究生, 主要研究方向为数据处理与分析、数据挖掘和通信网络故障。



刘立芳, 教授, 博士, 主要研究方向为数据处理与智能计算。主持国家自然科学基金项目、陕西省自然科学基金项目等国家和省部级项目 5 项。发表学术论文 40 篇。



齐小刚, 教授, 博士生导师, 博士, 主要研究方向为复杂系统建模与仿真、网络算法设计与应用。主持国家自然科学基金项目、十三五预研项目等国家和省部级项目 20 余项。获授权专利 19 项, 软件著作权 4 项。发表学术论文 100 余篇。