



# (12) 发明专利

(10) 授权公告号 CN 109559059 B

(45) 授权公告日 2022. 12. 27

(21) 申请号 201811542570.8

CN 102214213 A, 2011.10.12

(22) 申请日 2018.12.17

CN 104364664 A, 2015.02.18

(65) 同一申请的已公布的文献号

US 2018357542 A1, 2018.12.13

申请公布号 CN 109559059 A

邵帅. 传输技术在信息通信工程中的应用. 《数字技术与应用》. 2016, (第5期),

(43) 申请公布日 2019.04.02

黄勇等. 基于决策树算法的工业光纤网络通信故障检测系统. 《现代电子技术》. 2014, 第37卷 (第20期),

(73) 专利权人 重庆邮电大学

地址 400065 重庆市南岸区南山街道崇文路2号

曹俊. 铸坯质量的多元统计分析及决策树规则研究. 《中国优秀博硕士学位论文全文数据库 (硕士) 工程科技I辑》. 2009, (第6期),

(72) 发明人 王进 邵帅 许景益 孙开伟

邓欣 陈乔松

Wang JIn等. A novel optical fiber

(74) 专利代理机构 重庆市恒信知识产权代理有限公司 50102

专利代理师 刘小红 陈栋梁

doped with semiconductor nano-particles. 《Chinese Journal of Lasers》. 2008, 第35卷 (第10期),

(51) Int. Cl.

G06Q 10/06 (2012.01)

G06Q 50/04 (2012.01)

G06K 9/62 (2022.01)

G. Weckman等. Assessing Wireless Network Dependability through Knowledge Extraction via Decision Trees. 《Third International Conference on Systems (icons 2008)》. 2008,

(56) 对比文件

CN 108072596 A, 2018.05.25

CN 107923898 A, 2018.04.17

审查员 范琳琳

权利要求书1页 说明书4页 附图2页

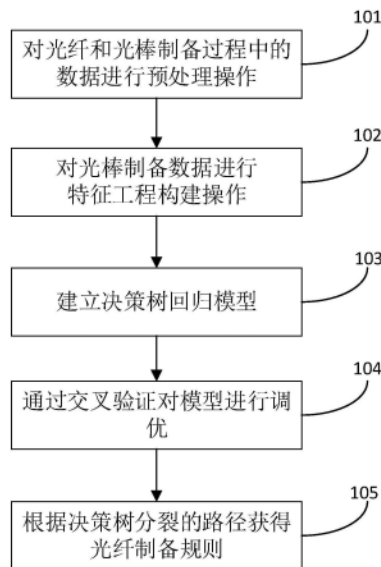
(54) 发明名称

一种基于回归决策树的光纤生产规则制定方法

(57) 摘要

本发明请求保护一种基于回归决策树的光纤生产规则制定方法,属于机器学习、大数据处理技术领域,具体包括步骤:101收集光纤生产过程中的光棒数据和光纤数据并对光棒数据和光纤数据进行预处理操作;102对光棒数据进行特征工程构建操作;103建立决策树回归模型;104通过交叉验证对模型进行调优;105根据决策树分裂的路径获得光纤生产规则。本发明主要是通过对光纤生产过程中的光棒数据和光纤数据进行预处理和分析提取特征,建立决策树回归模型,并对模型进行优化,通过决策树分裂的路径获取光纤生产的规则。

CN 109559059 B



1. 一种基于回归决策树的光纤生产规则制定方法,其特征在于,包括以下步骤:

101. 收集光纤生产过程中光棒的测量数据形成光棒特征数据,收集光纤的测量数据形成光纤特征数据,并对光棒特征数据和光纤特征数据进行噪声处理和缺失值处理预操作;

102. 链接光棒特征数据和光纤特征数据,对光棒特征数据进行特征工程构建操作;

103. 建立决策回归树模型;用光棒特征数据作为模型的特征,光纤特征数据的某个标签数据作为模型的标签,建立回归决策树模型;

104. 通过参数调节和交叉验证来对回归决策树模型进行调优,选择模型中平均绝对误差值低的模型作为最终的模型;

105. 根据回归决策树模型的分裂路径,获取多条由光棒特征和光纤标签组成的多条光纤生产规则;

所述步骤101收集光纤生产过程中光棒的测量数据形成光棒特征数据,收集光纤的测量数据形成光纤特征数据,并对光棒特征数据和光纤特征数据进行预处理操作:根据光棒特征数据和光纤特征数据的描述以及物理解释进行如下处理:

光棒特征数据和光纤特征数据中存在光纤小盘筛选长度为负值,占比值大于1的异常值,对于这些异常值进行去除行数据的操作;

如果特征值出现缺失,那么就使用热卡填充法寻找与它最相似的数据值进行补充;

所述热卡填充法是指对于一个包含缺失值的行数据,热卡填充法在完整数据中找到与它余弦相似度最高的行数据,然后用该行数据对应的值来对缺失值进行填充;

所述步骤102对光棒特征数据进行特征工程构建操作,具体包括:过滤光棒特征数据中小盘编号和日期字段的数据,剩余数据作为模型的特征;

所述步骤103建立决策回归树模型具体包括:用光棒特征数据作为模型的特征,光纤标签数据的某个标签数据作为模型的标签,建立回归决策树模型;

所述步骤104是采用十折交叉验证法去验证回归决策树模型,验证步骤包括:将数据集随机分成10份,轮流将其中9份作为训练数据,1份作为验证数据,进行验证模型,每次验证都会得出相应的平均绝对误差值,10次的平均绝对误差值的平均值作为模型的输出,通过此输出来验证模型和调节模型参数;

所述105. 根据回归决策树模型的分裂路径,获取多条由光棒特征和光纤标签组成的多条光纤生产规则,具体包括:一棵回归决策树包含一个根节点,多个内部节点和多个叶子节点,叶子节点对应决策结果,其它每个节点则对应一个属性测试,每个节点包含的数据集合根据属性测试的结果被划分到子节点中,从根节点到叶子节点的分裂路径对应一个判定路径,根据建立的回归决策树模型,此模型的每条分裂路径即为光纤生产的规则。

## 一种基于回归决策树的光纤生产规则制定方法

### 技术领域

[0001] 本发明属于机器学习、大数据处理技术领域,尤其基于回归决策树的光纤生产规则制定方法。

### 背景技术

[0002] 光纤产业是我国国民经济的重要支柱。近些年,随着通信和移动互联网的迅速发展,国内的光纤需求量日益增加,光纤产业在生产能力和技术实力上得到了快速提高,目前正朝着产业信息化,智能化和数据化发展。在光纤生产过程中,光纤生产专家根据光棒的测量参数来判断预测光纤的某些参数,而这些参数直接影响着光纤的质量。但是每个生产专家的判断标准是不一致,并且生产量变大、制作工艺有差距等因素,导致专家规则难以固化,难以形成一套完成的生产规则系统。本发明在传统光纤生产过程的基础上,一种基于回归决策树的光纤生产规则制定方法,数据预处理操作能有效的剔除数据中的冗余信息,特征选择能提炼出特征中的有效信息,既能提高算法的效率,还能提高算法的精准度。本发明采用回归决策树的机器学习方法,通过交叉验证法不断的优化回归决策树模型,根据决策树分裂路径获取准确的光纤生产规则。

### 发明内容

[0003] 本发明旨在解决以上现有技术的问题。提出了一种准确地获取光纤生产中的规则,并对这些规则进行固化,进而有效地提高光纤生产水平的基于回归决策树的光纤生产规则制定方法。本发明的技术方案如下:

[0004] 一种基于回归决策树的光纤生产规则制定方法,其包括以下步骤:

[0005] 101.收集光纤生产过程中光棒的测量数据形成光棒特征表,收集光纤的测量数据形成光纤表,并对光棒特征数据和光纤数据进行噪声处理和缺失值处理预操作;

[0006] 102.102.链接光棒特征数据和光纤数据,对光棒特征数据进行特征工程操作;

[0007] 103.建立决策回归树模型;用光棒特征数据作为模型的特征,光纤标签数据的某个标签数据作为模型的标签,建立回归决策树模型,

[0008] 104.通过参数调节和交叉验证来对回归决策树模型进行调优,选择模型中平均绝对误差值低的模型作为最终的模型;

[0009] 105.根据回归决策树模型的分裂路径,获取多条由光棒特征和光纤标签组成的多条光纤生产规则。

[0010] 进一步的,所述步骤101收集光纤生产过程中光棒的测量数据形成光棒特征表,收集光纤的测量数据形成光纤表,并对光棒数据和光纤数据进行预处理操作:根据光棒特征表、光纤表的描述以及物理理解进行如下处理:

[0011] ①光棒数据和光纤数据中存在光纤小盘筛选长度为负值,占比值大于1的异常值,对于这些异常值进行去除该行数据的操作;

[0012] ②如果某些特征值出现缺失的情况,那么就使用热卡填充法寻找与它最相似的数

据值进行补充。

[0013] 进一步的,所述热卡填充法是指对于一个包含缺失值的行数据,热卡填充法在完整数据中找到与它余弦相似度最高的行数据,然后用该行数据对应的值来对缺失值进行填充。

[0014] 进一步的,所述步骤102对光棒数据进行特征工程构建操作,具体包括:过滤光棒数据中小盘编号和日期字段的数据,剩余数据作为模型的特征。

[0015] 进一步的,所述步骤103建立决策回归树模型具体包括:用光棒特征数据作为模型的特征,光纤标签数据的某个标签数据作为模型的标签,建立回归决策树模型。

[0016] 进一步的,所述步骤104是采用十折交叉验证法去验证回归决策树模型,验证步骤包括:将数据集随机分成10份,轮流将其中9份作为训练数据,1份作为验证数据,进行验证模型。每次验证都会得出相应的平均绝对误差值。10次的平均绝对误差值的平均值作为模型的输出,通过此输出来验证模型和调节模型参数。

[0017] 进一步的,所述步骤105根据回归决策树模型的分裂路径,获取多条由光棒特征和光纤标签组成的多条光纤生产规则,具体包括:一棵回归决策树包含一个根节点,多个内部节点和多个叶子节点。叶子节点对应决策结果,其它每个节点则对应一个属性测试。每个节点包含的数据集合根据属性测试的结果被划分到子节点中。从根节点到叶子节点的分裂路径对应一个判定路径。根据建立的回归决策树模型,此模型的每条分裂路径即为光纤生产的规则。

[0018] 本发明的优点及有益效果如下:

[0019] 本发明将传统的光纤生产和机器学习,大数据等技术相结合,对光纤生产人员发现新生产规则,固化旧规则有辅助作用,进而优化光纤生产工艺,提高光线生产效能。

## 附图说明

[0020] 图1是本发明提供优选实施例提供基于回归决策树的光纤生产规则制定方法的流程图;

[0021] 图2为本发明实施例一提供基于回归决策树的光纤生产规则制定方法中回归决策树模型的分裂图;

## 具体实施方式

[0022] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、详细地描述。所描述的实施例仅仅是本发明的一部分实施例。

[0023] 本发明解决上述技术问题的技术方案是:

[0024] 参考图1,图1为本发明实施例一提供回归决策树的光纤生产规则制定方法的流程图,具体包括:

[0025] 101.收集光棒和光纤生产过程中的数据并对生产数据进行预处理操作:收集光纤生产过程中光棒的测量数据形成光棒特征表,收集光纤的测量数据形成光纤表。

[0026] 光棒特征表的具体特征有:日期、芯棒编号、小盘编号、测量人、光纤总长度、fiber\_step\_lens\_ratio、预计soot\_VAD沉积、实际soot\_VAD沉积、包层直径上\_VAD沉积、包层直径中\_VAD沉积、包层直径下\_VAD沉积、沉积实际长度\_VAD烧结、烧结芯棒长度\_VAD烧

结、长度\_PK2600芯棒、位置\_PK2600芯棒、DeltaPlus检验值\_PK2600芯棒、DeltaMinus检验值\_PK2600芯棒、DeltaTotal检验值\_PK2600芯棒、OD\_Core检验值\_PK2600芯棒、OD\_Clad检验值\_PK2600芯棒、CoreSlope检验值\_PK2600芯棒、D\_d检验值\_PK2600芯棒、Non\_core检验值\_PK2600芯棒、Non\_clad检验值\_PK2600芯棒、A\_esi检验值\_PK2600芯棒、Index1检验值\_PK2600芯棒、Index2检验值\_PK2600芯棒、实际有效长度\_拉伸、目标直径\_拉伸、拉伸长度\_拉伸、D1处直径\_拉伸、D2处直径\_拉伸、D3处直径\_拉伸、D4处直径\_拉伸、D5处直径\_拉伸、分切后长度\_粗切、原芯棒长度\_粗切、精切前长度\_精切、精切后长度\_精切、最大跳动值\_OVD跳动测试、总长\_OVD成品检、假芯棒长度\_OVD成品检、有效长度\_OVD成品检、净重\_OVD成品检、有效预拉长度\_OVD成品检、计费长度\_OVD成品检、计费重量\_OVD成品检、计费预拉长度\_OVD成品检、Bottom\_B\_OVD成品检、Top\_A\_OVD成品检、Cut\_Off\_OVD成品检、最大直径\_OVD成品检、最小直径\_OVD成品检、平均直径\_OVD成品检、密度\_密度测试、重量\_密度测试、沉积速率\_密度测试、Soot直径\_密度测试、有效长度\_密度测试、重量误差\_密度测试。

[0027] 光纤的测量数据具体有：日期、小盘编号、小盘筛选长度、Cladding Dia.、Cladding Cir.、Fiber Dia.、OCE、Fiber Cir.、ECC、芯不圆度、Primary coating Dia.、PCE、Core Dia.、测试长度、Att.-1310nm、1310nmA端衰减、1310nmB端衰减、Att.-1550nm、1550nmA端衰减、1550nmB端衰减、1310端差、1550端差、1310nm衰减不连续性、1550nm衰减不连续性、1310nm衰减不均匀性、1550nm衰减不均匀性、Att.-1383nm、Att.-1625nm、1310MFD、1550MFD、 $\lambda_c$ 、1383谱衰减、1625谱衰减、1285-1330nm范围内最大衰减与1310nm相比、1525-1575nm范围内最大衰减与1550nm相比、1310谱衰减、1550谱衰减、1460谱衰减、宏弯衰减/弯曲半径7.5mm圈数1(1550nm)、宏弯衰减/弯曲半径7.5mm圈数1(1625nm)、宏弯衰减/弯曲半径10mm圈数1(1550nm)、宏弯衰减/弯曲半径10mm圈数1(1625nm)、宏弯衰减/弯曲半径15mm圈数10(1550nm)、宏弯衰减/弯曲半径15mm圈数10(1625nm)、Curl.、Zero DML.、Slope zero DML.、1285-1339nm色散、1271-1360nm色散、Dispersion 1550nm、Dispersion 1625nm、1530-1565nm色散、1565-1625nm色散、PMD、光纤等级

[0028] 数据预处理包括对光棒特征数据和光纤数据的处理，根据两个数据表的描述以及物理解释进行如下处理：

[0029] ①由于光棒数据和光纤数据中存在一些异常值，比如小盘筛选长度为负，fiber\_step\_lens\_ratio大于1等异常值进行清洗处理，即去除该行数据；

[0030] ②由于光棒数据和光纤数据中某些特征存在缺失现象，针对数值型特征和枚举型特征分别做不同的处理，如果某些数值型特征出现缺失，比如1310MFD中会出现null值的情况，那么就用当前的中位数补充；光纤等级等枚举型特征用区别于数据集中现有的枚举值填充，如“未知”；

[0031] 102. 光棒数据进行特征工程构建操作：棒数据中无效特征是指在光棒特征表和光纤标签表需要通过某些字段将光棒特征数据和光纤标签数据连接在一起，但是这些字段对应的特征对于下文所述的回归决策树是无意义的，故需去掉。

[0032] 103. 建立决策树回归模型：建立回归决策树模型是指用光棒特征数据作为模型的特征，光纤标签数据的某个标签数据作为模型的标签，建立回归决策树模型，设置树的最大深度为4。

[0033] 104. 通过交叉验证对模型进行调优：用十折交叉验证发去验证回归决策树模型，

选择模型中MAE值低的模型作为最终的模型。

[0034] 105. 根据回归决策树分裂的路径获得光纤生产规则:根据回归决策树模型的分裂路径,可以获取多条由光棒特征和光纤标签组成的多条光纤生产规则。

[0035] 参考图2,图2为回归决策树的分裂图。当要分析光纤指标截止波长( $\lambda_c$ )和光棒特征之前的规则情况时,就可以用回归决策树去建立模型,形成多条分裂路径,每条路径就可以看作是一条关于 $\lambda_c$ 的生产规则。如最右边的分裂路径,可以得到一条生产规则为:

[0036] {计费长度\_OVD成品检>1430and fiber\_step\_lens\_ratio>0.052}= $\lambda_c$ 均值1238.1949/方差644.409。按此方法就可以获得多条关于光纤的生产规则。

[0037] 以上这些实施例应理解为仅用于说明本发明而不用于限制本发明的保护范围。在阅读了本发明的记载的内容之后,技术人员可以对本发明作各种改动或修改,这些等效变化和修饰同样落入本发明权利要求所限定的范围。

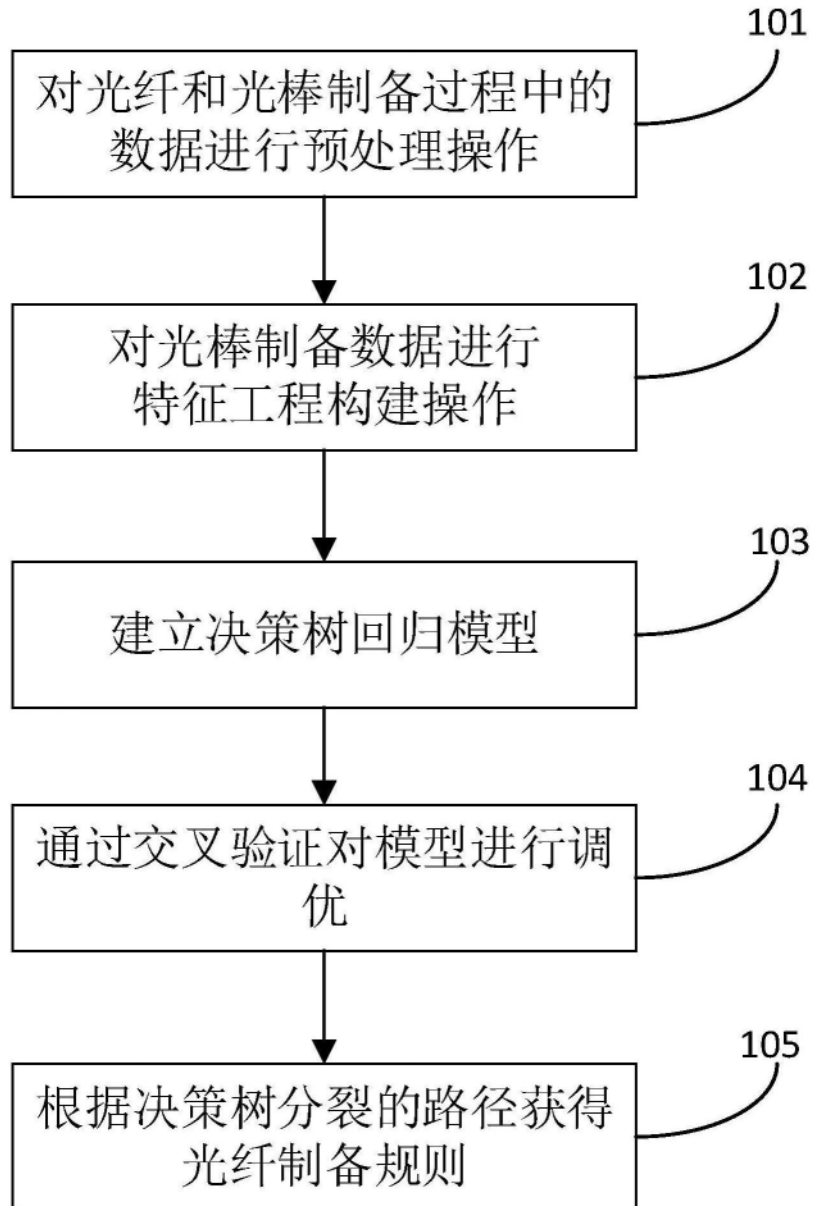


图1

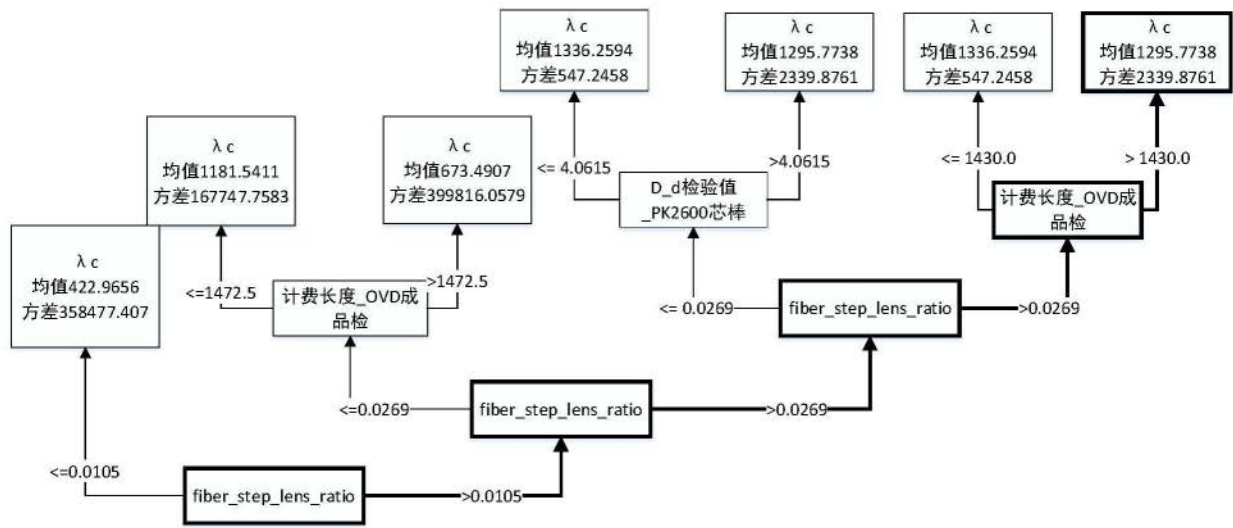


图2