

ID3 决策树算法在汽车检测中的分析及应用

邹黄刚, 赵景顺, 李彬

(上海理工大学机械工程学院, 上海 200093)

摘要: 针对现有的汽车检测方案, 存在耗时大、成本高且可实施性不强等缺陷, 现利用 ID3 决策树算法来设计一种新型汽车故障检查方案, 查找出何种因素引起的汽车故障, 使驾驶员自身具备故障检查能力, 并做出相应的预检修。通过将原始数据进行预处理后, 再运用 ID3 决策树算法挖掘初加工的原始数据, 建立起该数据下的决策数模型。利用刚建立的模型, 找出影响汽车发生故障的因素, 从而协助修理人员或驾驶员检修汽车, 提示维修和检测效率, 大大节约时间与成本。结果表明: 汽车最容易发生的故障是发动机故障, 影响其最大的因素是汽车消耗品质量差异。

关键词: 轿车; 汽车故障; 故障分析; ID3 决策树算法

中图分类号: TK0~TK91 文献标识码: A 文章编号: 1006—7973 (2022) 11-0153-03

汽车是由上万块零部件所组成, 因具备零件数量多、测试难、耗时多等特点^[1], 很难做到全方位检测; 并且大多数司机缺乏专业知识, 对汽车的故障很难做出准确的判断。所以, 对于检测人员或司机而言, 都急需一套快速而有效的检测方案, 指导其检修具有重要的意义。

本文以奔驰汽车为对象, 首先确定基于 ID3 决策树算法^[2]的汽车故障检修方案; 其次进行数据相关的采集, 这里数据信息采集首先确定好研究对象和目标, 选择合适的模型完成数据信息采集; 然后利用 ID3 算法的相关原理来对数据分析做出预处理以便实施选择的数据挖掘算法^[3], 最后对结果加以分析^[4], 其流程图如图 1 所示。



图 1 数据处理流程

1 汽车检测方案

1.1 现有的汽车检测方案

①鉴定故障: 通过故障信息, 判断出汽车何种故障。②初步确定好故障类别后, 通过询问车主或查看汽车来进一步收集故障信息。③收集完信息后, 进一步评估故障的危害有多大。④查明故障后, 按照一般的逻辑顺序进一步试验。⑤校正故障, 逐个排除故障原因。⑥重新逐个检测系统是否正常。如图 2 所示, 这六步检测^[5]虽

行布置。

(3) 随着横撑刚度的增大, 全桥稳定系数增大, 并趋于一个临界值。应根据实际情况来选择合适的刚度, 来保证桥梁稳定系数高的同时提高经济效益。

参考文献:

- [1] 朱贤明. 钢箱系杆拱桥顶推施工仿真计算及局部应力分析 [D]. 长安大学, 2018.
- [2] 唐细彪. 刚架系杆拱桥拱墩梁结合部结构受力分析及优化 [J]. 交通科技, 2020(05):21-24+34.
- [3] 尚晋. 中承式系杆拱桥施工控制关键技术 [J]. 内蒙古公路与运输, 2020(02):6-10.DOI:10.19332/j.cnki.1005-0574.2020.02.002.

[4] 李玲瑶, 陈政清, 葛耀君. 钢管混凝土拱桥拱肋横撑对动力和稳定性的影响 [J]. 公路交通科技, 2008(03):70-74+78.

[5] 董锐, 陈亚钊, 郑穆然, 黄福云, 陈宝春. 带 L 形横撑的大跨 CFST 枝式拱桥稳定性分析 [J]. 土木工程学报, 2020, 53(05):89-99+128.DOI:10.15951/j.tmgcxb.20200211.0001.

[6] 王海城, 黄己伟, 刘丹. 横撑系统对系杆拱桥动力特性的影响 [J]. 湖北理工学院学报, 2019, 35(05):36-39+72.

[7] 解瑶. 横撑布置位置对拱桥横向稳定性影响研究 [J]. 四川水泥, 2019(03):276.

[8] 杨吉新, 梁亚兰, 余越, 陈一赫, 杨蒋鹤立. 横撑对大跨度钢管砼拱桥稳定性的影响分析 [J]. 公路与汽运, 2018(05):118-120.

然能做到对汽车故障的全方位排查，但无疑会增加检修人员工作量，耗费大量的时间与成本，并且也不会给日常检修提供任何参考。

1.2 算法下的汽车故障检修方案

采用 ID3 决策树算法的汽车故障检测方案设计过程，如图 3 所示，第一层是视图层，即用于人机交互使用。第二层是业务层，即实现逻辑控制。第三层是 Model 层，负责有关软件功能与数据库系统和文件相互之间的数据访问。使用系统的用户包含司机和检修人员。

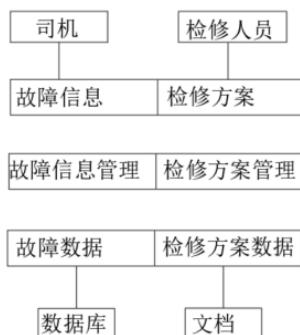


图 2 汽车故障检修方案流程

2 ID3 算法原理

ID3 算法用于预测分析，通过建立决策树，得出相关的检修方案。构造好的决策树关键在于选择好划分属性。决策树预测精度取决于所构建决策树的规模，一般而言，规模越小的，则其预测精度就越高。但想要构造出合理且精度高的决策树，关键在于选择恰当的测试属性，而测试属性的选择关键在于子集的不纯度度量方法查找，即所求的信息增益。

假设某个信息源集^[6]，标记为 S_i ， i 取 $0 \sim n$ ；其自信息量为 $I(S_i)$ ，则样本中的信息熵可以(1)式可得，如(1)式所示：

$$H(X) = -\sum_{i=1}^n p(S_i) \log p(S_i) \quad (1)$$

式中， n 为所有的信息源， $p(S_i)$ 为 S_i 的概率。

由前面得出目标信息熵后，再根据信息增益得出最优的测试属性 a ，通过信息增益的大小来确定决策树节点，依次构建决策树。样本各属性的信息增益运算公式如式(2)所示：

$$I(X; a) = H(X) - H(X|a) \quad (2)$$

其中：

$$H(X|a) = \sum_j P(a=a_j) H(X_j)$$

$$H(X_j) = \sum_j P(C_j | a_j) \log P(C_j | a_j)$$

式中， a 为样本 X 的属性； $H(X)$ 为信息熵； C_i 是训练实例； a_j 是第 j 个属性的概率。

3 ID3 决策树算法检修方案应用

3.1 汽车常见故障

目前汽车常见故障主要包括：发动机总成故障、传动系故障、转向系故障、制动系故障、行走系故障、车身常见故障、电气系故障等。引起以上故障的主要原因有：①自身存在着易损零件；②零件本身质量差异；③汽车消耗品质量差异；④汽车使用环境影响；⑤驾驶技术和驾驶方法的影响；⑥汽车故障诊断技术和维修技术的影响。由此可见，车辆故障现象普遍地产生于车辆的生产、使用、养护和检修工作的整个过程，对每一个环节都应十分注意，才能使车辆在运用工作过程中避免发生交通事故。属性 A 表示自身是否存在着易损零部件；属性 B 表示零件自身质量差异；属性 C 表示汽车消耗品质量差异；属性 D 表示驾驶技术和驾驶方法的因素；属性 E 表示车辆诊断技能和修理技能的影响；属性 F 表示汽车使用环境。表中“1”表示“是”，“0”表示“否”。

3.2 ID3 决策树的应用

由于汽车的型号特别多，每种车发生故障的形式也不同。本文以奔驰车为例，通过向大众发送询问调查报告的方式来统计出现以上七种故障的奔驰汽车，从而估计一般奔驰车会出现何种故障或者影响其出现故障的主要因素。投出去 60 份调查报告，实际参与人数返回报告 54 份，选取其中符合条件的调查报告 30 份来作为统计数据，进一步说明 ID3 算法是如何进行的。

汽车在使用过程中的故障有很多，现以汽车总成故障为训练集。数据样本是 30 条，故障的有 20 条，没有故障的 10 条。所以，根据(1)式，得故障信息熵是：

$$H(X) = 0.918$$

然后根据式(1)、式(2)分别计算各属性的信息熵和信息增益。同理可得影响因素初始时刻信息熵为：

$$H(X) = 0.882$$

再根据式(1)、式(2)分别计算各个影响因素的信息熵和信息增益。

从后期数据处理中可以看出，其“制动系故障”属

性信息增益最大，所以该属性作为决策树上层的第一个节点，随后各属性增益依次作为下一层节点。由此可知汽车在使用过程中最常出现的是制动系故障。影响因素表中，其“汽车消耗品质量差异”属性信息增益最大，所以该属性作为决策树上层的第一个节点，由此可知影响制动性故障最大的因素是汽车消耗品质量差异。同时考虑各属性之间的联系，再通过分支得到相应节点，最后得出“车辆故障”最常发生的是“制动系故障”，而“汽车消耗品质量差异”是此故障发生最大影响因素。

3.3 ID3 决策树算法的结果分析

由上述训练样本结论可知，面对汽车中的各类故障问题，通过统计相应的信息增益，得到样本中增益最大的属性，即能够得到区分度较大的特征属性及其属性值，从而明确汽车常发故障及其影响因素。ID3 决策树算法在以信息增益最高的那个属性视为一个节点的开始，随后进行节点的分割。每新建一个节点就会其延长一节，长此以往，一棵完整且合理的决策树就构建完成。而往往那些视为开始节点的属性也正是某些反常或故障现象的出现原因。当汽车设备异常或故障发生时，经过调 ID3 决策树算法，可得到产生影响该非正常或故障发生的主要原因，而后技术人员先对该原因加以检测。如是由该原因引起的，则对其提供相应保修；反之，对其他因素加以检测。该方案设计与对全部原因逐个加以检测的传统方法比较，更具科学化、优化和快速性，在相当程度上防止了某种非正常发展成故障现象，进而减少对牵引供电系统运行的负面影响。

4 结论

本文所提及的 ID3 算法旨在简化汽车检修模式，从而以最快的方式帮助司机或检修人员做出检修方案，避免不必要的浪费。利用 ID3 决策树设计的决策树，通过决策树上的节点找出汽车上的哪类故障以及出现此类故障的原因，进而帮助司机或检修人员快速做出较为合理的检修方案。以汽车各总成故障的数据作为训练集，对 ID3 决策树算法建立决策树流程加以分析计算，从而知晓其运作机理。

参考文献：

- [1] 王健, 庄继德编著. 汽车零部件发展之道 [M]. 北京: 国防工业出版社 .2013.
- [2] 杜威铭, 冉羽. 决策树 ID3 算法研究 [J]. 科技视界, 2018(11): 145–146.Du Weiming, Ran Yu. Research on ID3 algorithm of decision tree[J]. Science & Technology Vision, 2018(11): 145–146.
- [3] 雷波. 数据挖掘中分类算法的研究与应用 [J]. 电脑知识与技术, 2019, 15(27): 232–233.Lei Bo. Research and application of classification algorithms in data mining[J]. Computer Knowledge and Technology, 2019, 15(27): 232–233.
- [4] 张树滑. 基于 ID3 算法的大学生成绩数据挖掘与体能分析系统设计 [J]. 现代电子技术, 2019, 42(5): 104–106. Zhang Shuhua. Design of college student performance data mining and physical fitness analysis system based on ID3 algorithm[J]. Modern Electronic Technology, 2019, 42(5): 104–106.
- [5] 邹喜红主编; 张志强, 陈哲明副主编; 肖茂, 谢云峰, 征小梅等参编; 黄泽好主审. 汽车故障诊断技术 [M]. 北京: 中国铁道出版社 , 2011.08.
- [6] 李爱国, 库向阳编著. 数据挖掘原理、算法及应用 [M]. 西安: 西安电子科技大学出版社 .2012.

