

DOI:10.19431/j.cnki.1673-0062.2020.02.004

朴素贝叶斯分类在仪表故障判断上的应用

周自强

(辽宁红沿河核电有限公司,辽宁 大连 116000)

摘要:为了探讨朴素贝叶斯分类在仪表故障判断领域的应用价值,通过将某核电厂压力表故障的历史信息进行分类汇总,将故障的判断转换成文本分类任务,结合朴素贝叶斯分类算法和自然语言处理建立故障的分类模型,实现对新增故障的准确判断。通过验证,朴素贝叶斯分类模型能够对新增故障进行判断分类。测试中需要进行校验类故障准确率能够达到 95% 以上,其他类故障准确率高于 70%。传统故障判断一般是由人来完成,通过贝叶斯分类模型实现对故障的判断,可减轻人员劳动强度,提高工厂维修自动化水平。

关键词:朴素贝叶斯;分类;仪表;故障判断

中图分类号:TM407 **文献标志码:**A **文章编号:**1673-0062(2020)02-0021-05

Application of Naive Bayesian Classification in Instruments Fault Judgment

ZHOU Ziqiang

(Liaoning Hongyanhe Nuclear Power Co. Ltd., Dalian, Liaoning 116000, China)

Abstract:In order to investigate the application value of Naive Bayesian classification in the field of instruments fault judgment. By classifying and summarizing the historical maintenance information of pressure gauge fault in a nuclear power plant, the fault judgment is converted into a text classification task. After that the naive Bayesian classification algorithm and natural language processing are used to establish a diagnosis model to achieve accurate judgment of the new faults. It is proven the model can realize the fault judgment function, the accuracy rate which needs calibration can reach above 95%, while the others are above 70%. Traditional faults judgment is generally completed by human beings, while the bayesian classification model can reduce the labor intensity of personnel and improve the level of maintenance automation.

key words:Naive Bayesian; classification; instruments; fault judgment

收稿日期:2019-11-01

作者简介:周自强(1987-),男,工程师,硕士,主要从事仪表维修及维修策略优化方面的研究。E-mail:770993079@qq.com

0 引言

在工业生产上故障判断是维修活动的基础,维修人员在进行故障判断时首先根据故障描述进行故障分类,之后根据故障类别制定维修方案,这种传统的故障判断方式有一定的局限性,故障判断的准确与否直接受维修人员知识水平、工作经验的影响。在工业生产中如能实现对故障的自动准确判断,逐步摆脱对人的依赖,将能够大幅提高生产效益和安全生产水平。通过将某核电厂压力表历史故障信息进行整理后发现,故障判断可以归结为对故障的文本分类问题^[1]。

目前比较常用的文本分类方法有决策树、神经网络、k-近邻算法和朴素贝叶斯算法等^[2-3]。朴素贝叶斯分类器是贝叶斯分类模型的一种,文中通过应用朴素贝叶斯分类算法和自然语言处理技术,构建出一种故障判断模型,实现对故障的快速判断^[4]。

1 贝叶斯决策论

假设有 N 种可能的类别,即 $\lambda = \{c_1, c_2, \dots, c_n\}$, α_i 是将标记为 c_i 的样本准确分类产生的收益。通过后验概率 $P(c_i | x)$ 能够计算出将样本 x 分类为 c_i 的期望收益

$$E(c_i | x) = \sum_{i=1}^N \alpha_i P(c_i | x) \quad (1)$$

假如要目标 $x \rightarrow \lambda$ 的收益最大,就需要制定对应的判断规则 h

$$R(h) = E_x[R(h(x) | x)] \quad (2)$$

如每个样本 x 均能使 $R(h(x) | x)$ 的收益均最大,则总体收益最大,即

$$h^*(x) = \operatorname{argmax}_{c \in \lambda} R(c | x) \quad (3)$$

为了简化收益的计算,可假设收益 α_i 为常数 1,此时总收益

$$E(c | x) = P(c | x) \quad (4)$$

于是收益最高的分类器为

$$h^*(x) = \operatorname{argmax}_{c \in \lambda} P(c | x) \quad (5)$$

选择样本 x 使后验概率 $P(c | x)$ 最大的类别标记。利用贝叶斯公式来估计后验概率 $P(c | x)$ 在实际应用时由于类条件概率 $P(x | c)$ 相互之间的关联而存在困难。而朴素贝叶斯分类采用了属性条件独立性假设,对于给定的类别,每个不同的属性独立影响分类结果^[5]。基于属性独立假设,有

$$P(c | x) = \frac{P(c)P(x | c)}{P(x)} = \frac{P(c)}{P(x)} \prod_{i=1}^d P(x_i | c) \quad (6)$$

对所有类别来说 $P(x)$ 相当于一个常数,在属性独立假设的基础上,结合式(5)、式(6)。可得到朴素贝叶斯分类器如下

$$h_{nb}(x) = \operatorname{argmax}_{c \in \lambda} P(c) \prod_{i=1}^d P(x_i | c) \quad (7)$$

简单来说朴素贝叶斯分类器的构建就是先将目标的数据集合进行分类并计算出各类的先验概率 $P(c)$,对于各个类 c 寻找其特征属性 x_i 并计算特征属性的条件概率 $P(x_i | c)$,在朴素贝叶斯分类中特征属性对分类结果的影响相互独立^[6-7],于是特征属性的条件概率可以直接相乘,最后通过找到各个类别 $h_{nb}(x)$ 最大的数值,从而判定待分类数据的所属类别 c_i 。

2 贝叶斯分类算法的建立

2.1 数据的收集及预处理

通过对某厂就地压力表故障数据的长期收集,整理了 710 条相关故障及对应处理措施的数据。通过对数据的逐条整理将其分成了包括校验、清理管线并校验、接头紧固、更换、清洁并紧固、增加阻尼液、管线相关、其他等 8 类,通过分类后可以方便的得到各类故障在本数据集中的先验概率,如表 1 所示。

表 1 处理措施分类及其先验概率

Table 1 Classification of treatment and prior probability

处理措施	数量	先验概率/%
校验	400	56.3
清理管线并校验	28	3.9
接头紧固	50	7.0
更换	53	7.5
管线相关类	31	4.4
清洁并紧固	46	6.5
增加阻尼液	48	6.8
其他	54	7.6

2.2 特征属性建立

为了得到数据集的特征属性,需要对故障描述进行碎片化处理,这个过程通过自然语言处理的分词过程来实现。本文使用 jieba 工具完成对故障描述的中文分词。jieba 是一个现在流行的 Python 开源中文分词器,使用者可以方便的添加

用户词典,在当下自然语言处理领域应用非常广泛。

由于每种故障对应的故障现象不同,通过对故障描述进行分析,可以找到对应故障的特征描述,如某仪表故障描述如下“某月某日工作人员在某区域进行巡检时发现某压力表指示超量程,无法读取系统压力”,其中“指示”、“超量程”对故障的判定贡献很大,而描述中其他词语如“工作人员”“巡检”等对故障判定没有贡献,特征描述就是寻找对故障判定有指向性的词语。特征词汇

的选取过程是通过自然语言处理将某类故障的描述集中在一起,组成对该故障现象描述能全覆盖的集合,之后对出现的特征词汇进行收集(特征词选取越多,对分类准确性的贡献越大)。特征词汇的收集时主要选取高频词汇,也存在指向性强但出现频率低的词汇。

通过分词工具对分类后的数据进行分词并进行筛选和停用词剔除,分别找出每类故障对应的高频词汇和指向性词汇,将各类故障特征词汇汇总后作为总体特征序列,如表 2 所示。

表 2 特征词汇频率统计(部分)
Table 2 The statistics of characteristic vocabulary frequency (partial)

特征词汇	校验	清理管线 并校验	接头紧固	更换	管线相关类	其他	增加阻尼液	清油并紧固
校验	135	0	0	0	0	0	0	0
检查	133	0	0	3	0	0	0	0
处理	113	23	34	1	1	23	25	37
示数	26	0	0	0	0	0	0	0
不准	25	0	0	0	0	0	0	0
显示	24	0	0	0	0	0	0	0
指针	20	0	0	0	0	0	0	0
波动	13	0	0	0	0	0	0	0
漂移	6	0	0	0	0	0	0	0
堵塞	0	28	0	0	0	0	0	0
管线	0	23	0	0	32	0	0	0
引压管	0	8	0	0	7	0	0	0
取样	0	2	0	0	0	0	0	0
渗水	0	0	22	0	0	0	0	0
漏气	0	0	3	0	0	0	0	0
法兰	0	0	10	0	0	0	0	0
活结	0	0	8	0	0	0	0	8
漏水	0	0	8	0	0	0	0	0

2.3 模型建立

对各类故障分类和特征序列建立后,可以方便的建立朴素贝叶斯分类器。各类故障的先验概率 $P(c)$ 如表 1 所示,不同特征属性的条件概率 $P(x_i|c)$ 可以通过特征词汇频率进行计算。模型通过 Python 程序来实现,程序中直接给出各类的数量,将不同类的特征序列做成一个字典,结构如 {‘校验’: {‘校验’:135, ‘检查’:133, ‘处理’:113, ‘示数’:26, ‘不准’:25, ‘显示’:24, ‘指针’:20, ‘波动’:13, ‘漂移’:6, ‘堵塞’:0, …… ‘补充’:0, ‘液体’:0, ‘补液’:0, ‘添加’:0, ‘缺少’:0, ‘抗震’:0, ‘泄露’:0 } ……}。基于先验

概率和条件概率结合公式(7)便可以得到朴素贝叶斯分类器的模型,也就是压力表故障判断模型。

故障的具体判断是应用贝叶斯模型进行分类的过程,其判断流程如图 1 所示。首先对故障的描述进行分割处理,提取待判断故障的特征序列;之后将特征序列带入贝叶斯分类器,计算各类别的概率;最后对各类别的概率进行比较,找出其中的最大值,最大值对应的类别就是判断结果。

2.4 结果测试

为了验证模型的准确性,通过选取模型数据集和历史及新增故障数据集两类进行验证测试。首先通过模型数据集进行测试,故障判断示例如

图 2 所示,测试中发现校验类缺陷的分类准确率极高,选取的 200 项测试分类准确率达到 100%,其他类别测试准确率偏低,最低的不到 50%。根据样本数据集的测试结果,增加了模型特征序列中的特征词,这些特征词来自占比低的 7 类缺陷描述。在增加特征词的同时,人为增加特征词的出现频率,对模型进行调整。

在此之后选取了某电厂 464 项压力表类故障(历史及新增)进行验证,测试结果如表 3 所示,从测试结果可以看出,测试的准确率基本能够达到 80% 以上。

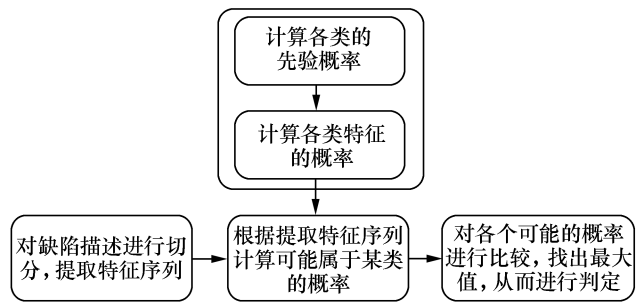


图 1 故障判断流程

Fig. 1 The process of troubleshooting

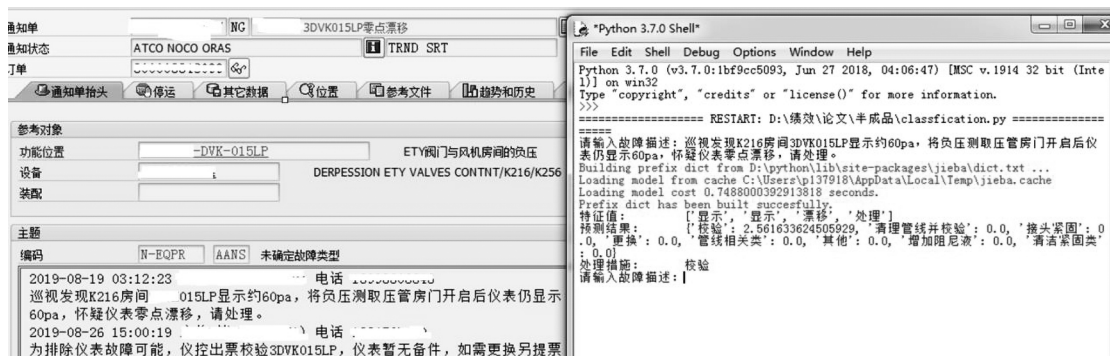


图 2 故障判断示例

Fig. 2 The example of troubleshoot

表 3 测试结果

Table 3 The test result

处理措施	数量	准确率/%
校验	230	99.57
清理管线并校验	24	83.33
接头紧固	40	92.50
更换	30	100.00
管线相关类	25	76.00
清洁并紧固	34	82.35
增加阻尼液	61	100.00
其他	20	80.00

通过对测试结果及模型原理分析,在创建模型之初选择样本容量大且全面的数据集对模型特征序列的选取很有帮助,在选取特征序列时要选取尽量多的词汇,这样能保证指向性高的特征词在特征序列之中。对于实际出现很少的缺陷,可以强制给出对应的特征序列并且设定特征词的出现频率,这样有助于提高分类的准确性。

3 结 论

故障判断方法基于概率统计,利用对历史故障数据的挖掘完成故障的判断。在进行模型准确性验证时,模型分类不能达到绝对准确,判断方法还需进一步的完善。在以后的工作中可以结合决策树、随机森林、支持向量机等分类算法,同时对现有模型增加进化功能来不断提高分类准确性。在普通生产型企业中,只要有常见仪表故障的历史记录信息,都可以采用该方法进行故障的分类判断。本文结合工厂数据,给出了朴素贝叶斯故障判断的实现流程和测试结果,对于实际系统故障诊断应用和相关领域的学术研究具有一定的参考价值。

参考文献:

- [1] 邹晓辉. 朴素贝叶斯算法在文本分类中的应用[J]. 数学技术与应用, 2017(12): 132-133.
- [2] 刘迪. 贝叶斯决策论[J]. 黑龙江科技信息, 2012(28): 71.

(下转第 33 页)

4 结 论

DCS网络故障时,由于不同处理器之间无法进行数据有效传输,将导致蒸发器水位控制信号部分丢失,无法实现DCS或BUP进行蒸发器水位控制。根据上文分析,针对蒸发器水位控制DCS部分失效模式,有如下3点建议:

1) B列蒸汽流量信号需进行改动,直接利用硬接线将B列信号传输至A列,一共需要增加三条模拟量控制电缆,将三个环路的B列蒸汽流量信号有效地传输至A列,并且,目前DCS机柜有足够的冗余通道可以实现这项工作,保障在DCS网络故障时蒸汽流量信号的有效性,防止调节阀波动。

2) 给水温度增益信号需进行改动,利用硬接线将给水温度增益信号分别送至每个蒸发器调节回路的DCS机柜。并且,目前DCS机柜有足够的冗余通道可以实现这项工作。从而保障给水温度变化对增益信号的有效输出,保证调节系统稳定性和灵敏性。

3) 蒸汽负荷信号不进行改动。如果要改动,则需要四个蒸汽负荷信号全部使用硬接线传输至每个蒸发器调节回路的DCS机柜。考虑到目前DCS机柜结构及卡件设置,无法通过冗余通道或卡件实现这项工作,且BUP部分优化后,DCS网络故障下,调节系统功能仍然完整。

对于BUP部分失效模式,认为增加质量位触

发的方式是十分可行的方法。该方法仅限软件修改,无需敷设电缆,工作量小,减少优化成本和优化后的维修工作量。但优化后,在DCS网络故障时,若误碰BUP工艺系统控制按钮会导致副调节阀误动的风险,需加强防人因失误意识。

参考文献:

- [1] 陈昱. 大亚湾核电站蒸汽发生器水位调节分析[J]. 科技视界, 2012(30): 261-264.
- [2] 郑广庆, 邢景伟, 赵星海, 等. 电厂汽管道水锤分析[J]. 能源研究与信息, 2012, 28(4): 237-241.
- [3] 侯长治, 陈洁, 崔瑶. 核电厂控制模式的工程设计研究[J]. 产业与科技论坛, 2017, 16(14): 66-68.
- [4] 石巍. 基于I/A Series的DCS网络设计[J]. 轻工科技, 2013, 29(7): 96-97; 144.
- [5] 谢虎. DCS系统网络故障分析与优化方案探讨[J]. 数字通信世界, 2017(8): 264.
- [6] KHALEEQ M T, WENPEN L, GUOSEN H. A novel computerized water level control system of PWR steam generator of nuclear power plant[J]. Journal of Shanghai university, 1998, 2(3): 226-227.
- [7] 杨小云. 浅谈提高DCS接线工艺质量[J]. 电工文摘, 2016(3): 68-70.
- [8] 温庆邦, 许斌亮, 尚艳菲, 等. 仪表失效后的处理[J]. 山东工业技术, 2014(22): 262.
- [9] 颜勇, 高俊成, 熊兴华. DCS核电站如何处理质量位的触发[J]. 山东工业技术, 2015: 60.

(责任编辑:扶文静)

(上接第24页)

- [3] 陈震, 潘尔顺. 基于退化数据的高可靠性产品贝叶斯分类决策[J]. 上海交通大学学报, 2017, 51(1): 76-77.
- [4] 彭艳斌, 郑志军, 于成波, 等. 基于分类器融合的自动化协商决策模型[J]. 上海交通大学学报, 2013, 47(4): 644-649.
- [5] 周志华. 机器学习[M]. 1版. 北京: 清华大学出版社, 2016: 147-155.

- [6] YANG Y X. An improved recommendation algorithm for micro-blog network advertisement[R]. Proceedings of 2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science. Wuxi, China: DCABES, 2018.
- [7] 王刚, 牛宏侠. 贝叶斯网络在计算机连锁系统故障诊断中的应用[J]. 电子世界, 2018(5): 174-175.

(责任编辑:周泉)