

基于机器学习的 P 公司总装质量预测方法

张云川, 倪静, 杨怡君, 李季, 李虎

上海理工大学 上海

【摘要】机器学习作为一个近期很热门的话题, 被越来越多的研究所应用。为了给 P 公司总装质量管理提供更好的技术支持, 提出了一种基于深度置信网络 (DBN) 的产品质量预测模型, 可通过产品装配历史数据的支持, 结合现场的变化情况, 判断出产品的合格率情况。以 P 公司某产品 2019 年的生产数据作为初始样本, 经影响因子提取、去除噪声和筛选稳定点等处理后, 形成最终样本集。利用 DBN 网络对样本数据进行特征提取, 建立 P 公司某产品质量预测模型。结果表明: 通过测试集验证, 基于参数调优的 DBN 模型能准确地预测状态参数, 拥有优良的稳定性。

【关键词】质量管理; 深度置信网络; 影响因子

Final assembly quality prediction method of P Company based on machine learning

Yunchuan Zhang, Jing Ni, Yijun Yang, Ji Li, Hu Li

Shanghai University of Science and Technology, Shanghai

【Abstract】As a hot topic recently, machine learning has been applied by more and more researchers. In order to provide better technical support for the quality management of the final assembly of P Company, a product quality prediction model based on deep confidence network (DBN) is proposed. The product qualification rate can be judged by the support of the product assembly history data and the change of the site. The 2019 production data of a product of Company P is taken as the initial sample, and the final sample set is formed after processing such as impact factor extraction, noise removal and stable point screening. The DBN network is used to extract the features of sample data, and a product quality prediction model of P Company is established. The results show that the DBN model based on parameter tuning can accurately predict the state parameters and has good stability through test set verification.

【Keywords】quality management, deep confidence network, influence factor

引言

近年来, 汽车召回, 三聚氰胺, 地沟油, 老坛酸菜等事件暴露出无论是食品还是制造行业, 产品质量问题都触目惊心。而网购, 公共服务, 医疗服务, 保险理赔等领域也经常曝出纠纷和服务投诉问题, 服务质量也不容乐观, 可见质量问题无处不在, 有很多问题需要解决。在当前社会快速发展的背景下, 处理质量问题手段也从解决慢慢变成预防。随着信息处理技术的不断发展, 各行各业已建立了很多计算机信息系统, 积累了大量的数据。为了使数据能够有效地支持组织的日常运作和决策, 要求数据可靠无误, 能够准确地反映现实世界的状况。数据是信息的载体, 好的数据质量使各种数据分析(如

OLAP 分析、数据挖掘等)能够得到有意义结果的基本条件。人们常常抱怨所谓的“数据丰富, 信息贫乏”, 其中一个原因是缺乏有效的数据分析技术, 而另一个重要原因则是数据质量不高, 如数据残缺不全、数据不一致、数据重复等□导致数据不能有效地被利用数据质量管理如同产品质量管理一样贯穿于数据生命周期的各个阶段但目前尚缺乏一个系统的思路。数据质量的研究由来已久, 涉及到统计学、人工智能、数据库等各个领域^[1]。P 公司拥有成套的自研 MES 系统, 也使用着 SAP 系统, 两套系统相互配合, 收集了非常多的生产数据, 可以运用于质量管理中。

1 田口方法质量模型

1.1 输出质量特性

田口方法中的输出质量特性共有三种。按照质量特性来源可以分为下游、中游、上游和源游质量特性。按照衡量尺度可以分为计量特性值和计数特性值。按照质量特性目标的趋向可以分为望目特性行, 望小特性和望大特性^[2]。P 公司总装产品主要为光刻机电源及冷却控制模组, 产品主要由客户设计, 公司需要按照图纸生产, 测试其电流, 电压, 通过后发货。故可定义为中游质量特性, 衡量尺度为计量特性, 目标值趋向为望目特性。

1.2 影响输出质量特性的因素

影响质量波动的因素很多, 按照 4M1E 可分类为人员的行为, 机器设备, 原材料, 方法以及环境。或者从设计的观点来看, 可分为两类。一类是可控因素, 是指那些在设计和制造中可以控制的因素, 例如零部件的几何尺寸、间隙, 所使用的原材料强度, 电阻的欧姆值等。这些因素又称为设计变量。另一类是不可控因素, 在设计中难以控制的因素。例如: 使用条件、生产环境、操作人员等。这些因素又称为噪声因素。P 公司总装部门负责的产品大多为代工产品, 不具有设计原因, 故本文影响质量特性的因素都集中在生产过程。

1.3 产品质量设计模型

按照田口质量模型, 如图 1, 产品质量由可控因素和不可控因素影响。

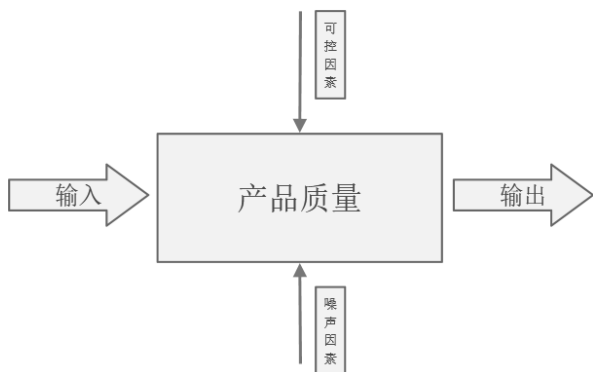


图 1 产品质量由可控因素和不可控因素影响

根据 P 公司总装部门实际情况, 产品质量的印象因素两者都有。比如客户产品的复杂程度、员工的熟练度、物料供应状况、作业指导书、ESD 环境等。由于总装过程多为小批量多品种, 复杂程度相差较大的产品, 所以影响因素可以确定为:

(1) 产品复杂程度, 根据产品来讲也就是 BOM

中紧固件数量占比。

(2) 员工熟练度, 根据能力矩阵可以了解一个员工的技术水平。

(3) 设备情况, 主要由维保及时率以及校准及时率组成

(4) 物料供给率, 由物料完备程度和供给及时率组成

(5) 装配复杂程度, 由页数和每页操作步骤组成

2 机器学习逻辑回归模型

2.1 逻辑回归的基本原理

逻辑回归本质上是一个 sigmoid 函数, 取值范围是 0 到 1, 如图 2 所示。

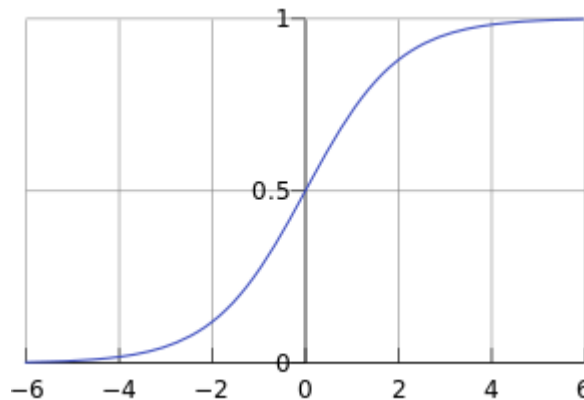


图 2 逻辑回归

产品的质量最终就是合格与不合格, 是一个二分类问题, 适合 sigmoid 函数。由于分成两类, 我们便让其中一类标签为 0, 另一类为 1。我们需要一个函数, 对于输入的每一组数据 $x^{(i)}$, 都能映射成 0~1 之间的数。并且如果函数值大于 0.5, 就判定属于 1, 否则属于 0。而且函数中需要待定参数, 通过利用样本训练, 使得这个参数能够对训练集中的数据有很准确的预测。

2.2 逻辑回归的应用方式

P 公司日常生产中所有产品都会生成序列号, 便于追溯, 所以拥有海量的产品质量数据。将生产过程中的数据也就是影响因素, 作为逻辑回归模型的输入, 历史产品质量数据作为标签, 不合格品为 0, 合格品为 1, 以此数据集可以训练出一套适用于 P 公司总装流程的质量预测模型。也就是通过生产过程中的变量, 来预测最终产品的质量。此方法可以方便项目团队在量产前, 通过小批量的样件过程,

了解大规模量产的合格率, 提前获知一些影响产品质量的风险点, 提前发现问题, 解决问题, 减少了公司量产过程再调节改善的成本^[3]。

3 机器学习在质量管理中的应用

3.1 影响因素的数据化

影响因素 1: 产品复杂程度

系统组装部门的产品以控制柜, 冷却系统, 大型电源系统为主。安装方式主要以螺丝, 铆钉, 螺母等紧固件为主。产品复杂程度可以用紧固件数量占比 BOM 总零件数量的比率来表述。

影响因素 2: 装配复杂程度

系统组装部门的装配指导书称为 PAD (Product Assembly Document), 每页一张实物照片, 辅以标准化的符号标记表示装配手法及要求。因为有标准化的格式, 所以 PAD 的页数即可表示装配复杂程度。^[4]

影响因素 3: 员工熟练度

系统组装部门产品全部都是人工组装, 那么员工的熟练程度可以由多次组装工时的稳定程度来描述。

影响因素 4: 物料完备程度

物料的完备程度直接影响能否完成制造, 数据来源与 SAP 中的 MRP 系统。

影响因素 5: 设备情况程度

产品的制造离不开制造线, P 公司制造线布置逻辑为按照工艺布置, 同一台设备会被不同产品的制造过程所使用。所有设备都通过 MES 系统中的 C&M 模块进行追踪, 在 Power Bi 中会有提取与整合。

3.2 机器学习预测的实现

使用 Anaconda 作为环境搭建平台, Python 作为编程语言, Visual studio Code 作为解释器, Pytorch 为基础架构^[5]。将数据化后的影响因素组成表单, 历史产品质量作为数据标签, 可以得到逻辑回归模型所需的数据输入表, 部分截取下表 1 所示。

在程序中将 ok 定义为 1, nok 定义为 0, 并且整体数据划分为训练组和验证组。使用基本的逻辑回归模型进行训练, 得到图 4 所示结果: 可以看出在 15 轮训练后模型基本达到 98% 预测准确率, 准确率迭代如表 2 所示。由此可见数据集被较为清晰地分为了 ok 与 nok 两类, 意味着该模型可以通过输入数据预测出结果。

表 1 数据输入表

5.4	3.9	1.7	0.4	ok
4.6	3.4	1.4	0.3	nok
5.0	3.4	1.5	0.2	ok
4.4	2.9	1.4	0.2	ok
4.9	3.1	1.5	0.1	nok

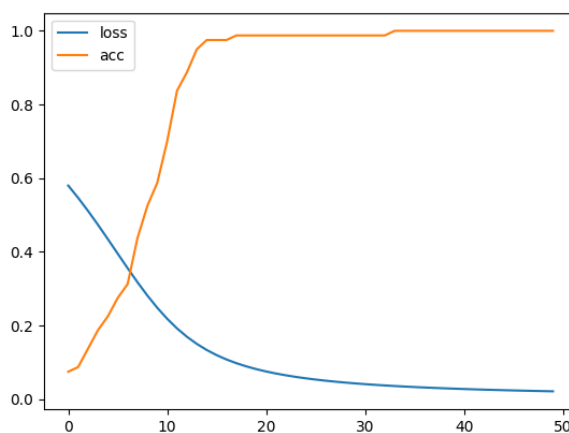


图 4 结果

表 2 准确率迭代

损失率	准确率
0.580041229724884	0.07500000298023224
0.5468625426292419	0.08749999850988388
0.5113701224327087	0.13750000298023224
0.47391408681869507	0.1875
0.4350321888923645	0.22499999403953552
0.3954528570175171	0.2750000059604645
0.3560629189014435	0.3125
0.31783413887023926	0.4375
0.28170865774154663	0.5249999761581421
0.2484693080186844	0.5874999761581421
0.21863500773906708	0.699999988079071
0.19241873919963837	0.8374999761581421
0.16975617408752441	0.887499988079071
0.15038299560546875	0.949999988079071
0.13392533361911774	0.9750000238418579
0.11997468769550323	0.9750000238418579
0.10813784599304199	0.9750000238418579
0.0980626791715622	0.987500011920929
0.08944748342037201	0.987500011920929
0.08204038441181183	0.987500011920929

4 总结

本文开展了使用机器学习逻辑回归模型进行产品质量预测的研究, 提出了一种基于生产数据和影响因子相结合的预测方法。于数据中分析质量问题, 尝试了将新进的机器学习与传统的田口模型相结合的应用方式。从大批量生产以来, 人们就不断探讨大数据的管理和应用的方式。真正的革命不是关于数据, 而是关于统计以及利用数据获得洞察方法所产生的价值。现在, 业界开始通过数据标注, 建立用于模式识别的人工神经网络训练和自学习, 实现基于机器学习的模式识别。根据世界知名的质量工程/可靠性工程/精益六西格玛专家, 美国韦恩州立大学终身教授杨凯的总结, 未来的产品开发制造将形成一个全面连通的生态圈, 其中人员、产品、机器、相关系统实现信息无障碍流通。物联网器、芯片等将植入硬件产品, 云计算、深度学习、人工智能、VR、Augmented Reality 等高新技术进入质量管理实用, 在风险与机遇并存中不断推进质量管理的深度和广度。在这个不断反复的过程中数据中的

问题逐步被发现解决从而使质量管理得到保证和提高。因此未来在制造业中通过大数据来提前预测、发现、识别产品质量问题会越来越普遍, 对于质量管理者也提出看了更高的要求, 需要掌握更多的数据分析及应用技能。

参考文献

- [1] 韩京宇, 徐立臻, 董逸生. 数据质量研究综述[J]. 计算机科学, 2008(02): 1-5+12.
- [2] 梁工谦, 刘德智, 陈洪根. 质量管理学[J]. 北京: 中国人民大学出版社, 2021.
- [3] 阿拉杜 (Alladoubaye N. 用于数据质量管理的鲁棒机器学习算法[D]. 哈尔滨工业大学, 2021.
- [4] 赵文浩, 邢香园, 王治, et al. 基于知识挖掘的航天产品装配工时定额和管理系统[J]. 计算机集成制造系统, 2021, 27(06): 1594-1604.
- [5] 何俊. 齿轮箱振动特性分析与智能故障诊断方法研究

[D]. 浙江大学, 2018.

收稿日期: 2022 年 9 月 18 日

出刊日期: 2022 年 11 月 28 日

引用本文: 张云川, 倪静, 杨怡君, 李季, 李虎, 基于机器学习的 P 公司总装质量预测方法[J]. 国际计算机科学进展, 2022, 2(4): 1-5.

DOI: 10.12208/j.aics.20220056

检索信息: RCCSE 权威核心学术期刊数据库、中国知网 (CNKI Scholar)、万方数据 (WANFANG DATA)、Google Scholar 等数据库收录期刊

版权声明: ©2022 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS