

# 面向微波组件工艺失效分析的大数据建模技术

徐榕青, 张晏铭, 王 辉, 李 扬, 庞 婷

(中国电子科技集团公司 第 29 研究所, 成都 610036)

**摘要:** 快速、准确定位微波组件生产过程中的质量问题, 对提升微波组件的质量可靠性、工艺稳定性以及生产效率具有重要作用。在微波组件工艺质量问题传统人工分析逻辑基础上, 通过对当前微波组件生产流程各环节的数据特点的挖掘分析, 提出了生产大数据与失效分析知识融合的建模方法, 并应用于工艺问题的辅助排故中; 首先, 基于微波组件工艺质量数据特征进行数据清洗得到故障关键数据, 作为大数据挖掘建模的基础数据; 其次, 从微波组件质量特征相似性的角度对不同微波组件进行聚类处理, 提升稀疏数据的信息密度; 最后, 采用大数据挖掘算法融合失效分析先验知识建立用于辅助排故的知识模型, 并基于样本数据对提出的建模方法进行了实例分析和模型的软件化部署, 验证了在微波组件工艺质量问题分析应用中的可行性。

**关键词:** 微波组件; 工艺质量; 失效分析; 知识融合

## Big Data Modeling Technology For Microwave Component Process Failure Analysis

Xu Rongqing, Zhang Yanming, Wang Hui, Li Yang, Pang Ting

(The 29th Research Institute of CETC, Chengdu 610036, China)

**Abstract:** It plays an important role in improving the quality reliability, process stability and production efficiency of microwave components that rapid and accurate positioning of the quality problems in the microwave components production process. Based on the traditional manual analysis logic of microwave components process quality problems, through the mining and analysis of the data characteristics of every link in the current microwave components production process, the modeling method of production big data and failure analysis knowledge fusion is put forward, and it is applied to the auxiliary troubleshooting of process problems. Firstly, based on the characteristics of microwave components process quality data, the key fault data is obtained through data cleaning, which serves as the basic data of big data mining modeling. Secondly, from the perspective of quality feature similarity of microwave components, cluster different microwave components to improve the information density of sparse data. Finally, the big data mining algorithm is used to fuse the prior knowledge of failure analysis to build the knowledge model for auxiliary troubleshooting. Based on the sample data, the case analysis and the software deployment of the model are carried out for the proposed modeling method, which verifies the feasibility of the application in the analysis of microwave component process quality problems.

**Keywords:** microwave component; process quality; failure analysis; knowledge fusion

## 0 引言

当前制造企业基于传统表层信息的报表分析、统计分析、趋势图查看等过程质量管控方式, 难以实现在数据类型混杂、关联耦合强、信息密度低、时间跨度大状态下的生产过程管控与问题分析, 数据中隐藏的规律和价值也难以有效发掘和利用。大数据挖掘模型不依赖于精确数学关系的特性, 适用于基于统计规律下的质量管控与分析, 通过模型的应用可降低对人员经验的依赖。当前标准化的数据挖掘工具和模型算法, 为质量大数据挖掘应用提供了技术基础, 企业可专注于产品生产过程分析方便快速地实现

应用验证<sup>[1-4]</sup>。面对生产过程中故障数据稀疏, 导致的大数据模型扩展性不强的问题, 可通过抽取设备、问题的特征标签, 采用聚类算法进行归类合并, 进而进行相似产品或问题处理专家知识的智能推荐, 有效提升了数据稀疏下所建模型的预测能力<sup>[5]</sup>。运用数据分析和机理分析融合的方法对数据进行处理, 可有效利用大数据的多尺度特性对装备生命周期进行分析预测、运行优化、预知维修, 保证了数据知识获取、运用的时效性<sup>[6]</sup>。通过数据特征分析、数据采集、处理、预测应用, 结合具体业务场景可建立基于大数据技术的应用信息系统, 实现质量管控与提升<sup>[7]</sup>。

针对新一代微波组件产品呈现出工艺更加复杂、返工返修难度更大、单件价格更高的特点, 业内鲜有采用大数据进行工艺质量分析的报道, 本文通过大数据挖掘方法对微波组件生产中的多维质量数据进行分析, 对深层信息和隐含规律的挖掘提取, 融合工艺失效分析先验知识建立大数据模型辅助工艺排故, 持续提升过程管控快速、准确响应能力, 促进工艺稳定性和产品质量提升。

收稿日期: 2020-07-27; 修回日期: 2020-08-11。

基金项目: 国家国防科技工业局项目(JSZL2018210B009)。

作者简介: 徐榕青(1977-), 男, 福建福州人, 硕士研究生, 高工, 主要从事微波集成技术方向的研究。

通讯作者: 张晏铭(1991-), 男, 四川乐山人, 硕士研究生, 工程师, 主要从事微波集成工艺及数字化技术方向的研究。

### 1 微波组件工艺失效分析流程

典型微波组件生产所采用的微组装机工艺, 如图 1 所示, 主要包括可制造性审查、工艺流程拟制、SOP 编制、计划下达、生产加工(图纸、工艺、机加件、元器件、材料、设备、人员等多要素耦合过程)、过程检验、调试测试(设计、制造与电性能多要素耦合)、终检、交付, 环节众多导致质量数据分析存在数据来源多、种类多、存储分散, 主要数据来源有:

- 1) 生产过程数据, 如设计规范、工艺规范、检验数据、调测数据;
- 2) 基础数据, 如产品基础数据、生产基础数据等只与产品、设备物料有关的基础数据;
- 3) 技术报告, 如归零报告、实验报告、总结报告等为处理问题、优化工艺开展技术研究形成的资料文档。

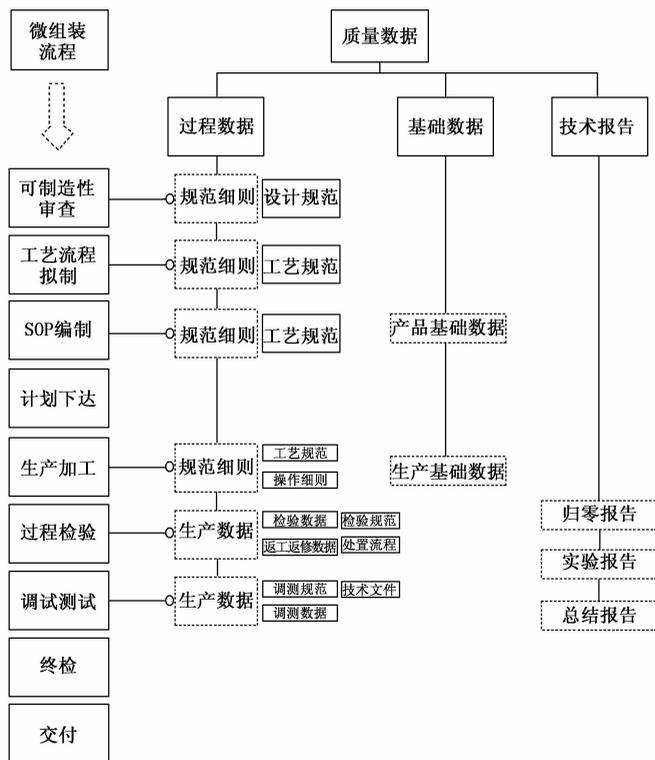


图 1 微波组件生产流程及质量相关数据

从内容角度, 各环节产生的数据主要分为 3 类: 生产相关、技术开发相关、规范相关, 数据间逻辑复杂交联, 且一般仅生产过程中的数据为数据库状态, 可基本满足采用数理统计等方法处理的要求。

为实现微组装机过程质量和技术状态管理, 当前对过程检验、返工返修等原始事务数据的应用模式局限于时序趋势监控、比率指标统计。当需要进一步深入分析工艺失效问题时, 主要通过讨论、生产过程数据分析、资料研究等复杂的流程, 由人工分析定位得到问题的原因和解决措施, 以 PFMEA 报告形式形成经验参考。

PFMEA 即工艺过程失效模式及影响分析, 如图 2 所

示, 其目的是假定产品设计满足要求的前提下, 针对产品在生产过程中每个工艺步骤可能发生的故障模式、原因及其对产品造成的所有影响, 按故障模式的风险优先数 (RPN) 值的大小, 对工艺薄弱环节制定改进措施, 并预测或跟踪采取改进措施后减少 RPN 值的有效性, 使 RPN 达到可接受的水平, 进而提高产品的质量和可靠性。

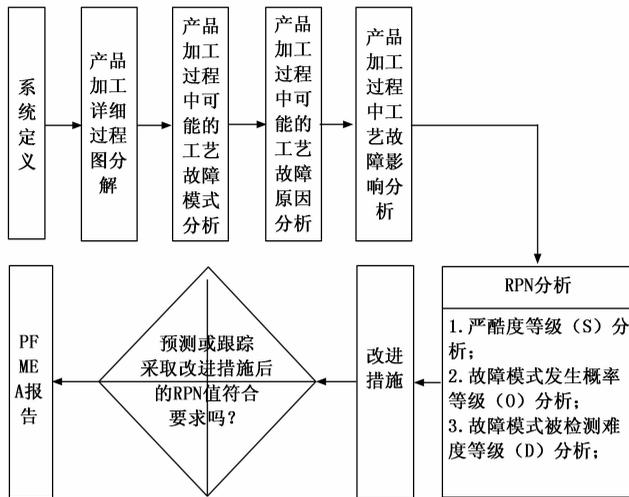


图 2 微波组件 PFMEA 流程

PFMEA 本身是工序强相关的, 微波组件 PFMEA 结合了生产过程的诸多维度, 如产品物料、工艺要求、设备要求等, 信息关联度高, 具有严密的逻辑性和理论支撑, 因此可作为领域先验知识指导各维度质量数据的提取, 包括工序、工步、工艺特征(过程功能要求)、失效模式(潜在失效模式)、失效原因(潜在失效要因)、失效后果(潜在失效后果)、处理措施(建议措施)等。

但由于微波组件具体结构、工艺参数、物料组合繁多, 导致数据分散、价值密度低, 仅依靠人工进行 PFMEA 分析难度较大, 时效性和覆盖度较差, 通过大数据挖掘可全面、准确把握微波组件生产过程各环节数据信息, 提升工艺质量问题分析处理有效性。

### 2 微波组件工艺质量数据建模

#### 2.1 质量数据特征

为便于大数据挖掘算法的处理, 建立工艺失效分析知识模型, 明确微波组件工艺质量数据具有如下基本类型和特点:

1) 数据维度多, 变量取值多。由于生产环节多, 导致产品工艺质量问题的因素存在于物料特性、工艺参数、设备参数、操作过程多个维度, 且各维度变量取值较多, 相应模型和方法也必须适用于多维度数据关联分析, 可理解数据中的规律;

2) 数据间关系不显性, 存在弱相关或强非线性的关系, 如故障分析通常需要经过多层级的模式一原因追溯, 较难通过简单的线性分析发现各层级要素之间的关系, 导致数据挖掘模型复杂;

3) 数据多以非结构化、离散型为主, 包括二元、标称、序数类型, 如问题产品是否需要返工(二元)、诱发问题的原因(标称)、问题的严重程度(序数)。

因此在进行建模分析之前结合微组装质量领域知识和 PFMEA 先验知识按照清洗、集成、变换、规约进行预处理将数据转换成标准的数据类型, 有利于提升模型的质量<sup>[8]</sup>。

再基于 PFMEA 建立用于描述失效分析的属性字段, 包括工序名、工序特征、工步名、产品信息、人员信息、失效后果、处理措施、失效原因、失效模式, 如图 3 所示。各属性字段类型均为离散标称型, 一般取值采用专业领域的标签化词语或序数。

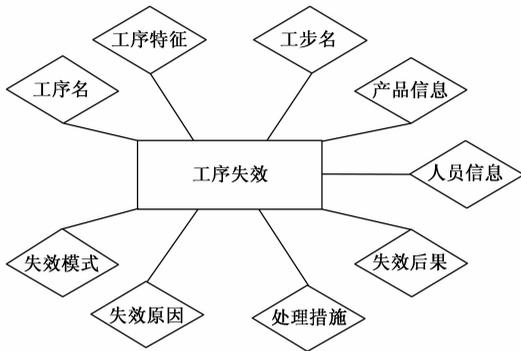


图 3 微组装质量数据分析属性字段

通过对历史工艺失效数据处理, 并采用标准属性字段和标签值进行清洗替换, 得到可用于大数据算法建模的规范数据, 如图 4 所示, “工序”属性字段包含粘接、共晶等标准微组装工序名作为字段值, “工序特征”属性字段包含一组工序具体信息的结构化词组作为字段值, 该字段与产品强相关, 相同工序名中一般包含不同的工序特征。基于此, 可对质量数据实现计数统计和数学运算处理。

工序名	失效模式	失效原因	严重程度
粘接	水汽超标	湿敏器件漏识别	1
共晶	焊盘氧化	烘烤时间过短	2
错装	爆米花效应	标识错误	...
钎焊	.....	.....	10
.....			

工序特征
陶瓷电路片, 可伐载板, 双组份胶, 温度, 时间 氧化铝芯片, 可代载板, 金锡合金, 温度, 时间

图 4 微组装质量数据各属性字段取值示例

### 2.2 质量特征聚类算法

微波组件产品之间从工序构成角度, 通常具有一定的相似性, 如 LTCC 基板构成的产品通常有应力开裂、器件粘接脱落的问题, 可以从工序构成、故障模式、故障原因等失效相关数据字段中首先提取相关性较高的字段作为聚类特征, 然后有多个不同的聚类特征构成特征向量对产品进行聚类, 当几种不同名称产品具有近似的聚类特征时, 则将这几种产品归为一类, 此类产品具有相似的质量特征。于是, 聚类模型输入输出数据格式如表 1 所示。

表 1 聚类分析输入输出数据格式

变量	输入				输出
	向量	向量	向量	向量	标称
名称	产品工序	工序特征	故障模式	故障原因	产品类别

聚类分析中 K 均值算法适用数据类型广泛<sup>[9]</sup>, 主要计算式如式 (1)、(2), 通过计算对象之间的邻近性和聚类的目标函数, 即凝聚度, 基于最大化簇中文本条目与簇质心的相似性并迭代计算可得最终聚类结果。总凝聚度, 如下所示:

$$TotalCohesion = \sum_{i=1}^K \sum_{x \in C_i} cosine(x, c_i) \quad (1)$$

$x$  是输入数据对象,  $C_i$  是第  $i$  个簇,  $c_i$  是簇  $C_i$  的质心, 是簇中数据的均值,  $K$  是簇的总数, 其中  $cosine\langle \cdot \rangle$  计算式如下, 表示两条数据向量之间的余弦:

$$cos(\theta) = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{\sqrt{(x_1^2 + x_2^2 + \dots + x_n^2)} \cdot \sqrt{(y_1^2 + y_2^2 + \dots + y_n^2)}} \quad (2)$$

对所有数据与各自簇质心的余弦进行求和计算, 通过算法优化簇质心找出总凝聚度最大的聚类分簇结果。

聚类分簇的有效性, 即好与差采用凝聚度进行评估, 主要评价指标为轮廓系数:

$$s_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (3)$$

对第  $i$  个对象, 计算它到簇中所有其他对象的平均距离, 记作  $a_i$ ; 对第  $i$  个对象和不包含该对象的任意簇, 计算该对象到给定簇中所有对象的平均距离。关于所有的簇, 找出最小值, 记作  $b_i$ 。

轮廓系数的值可取范围在 -1 到 1 之间, 通常情况下为 0 到 1, 且越接近 1 则说明聚类效果越好。

通过对产品进行分簇聚类, 可有效利用产品工艺特性及质量问题的相似性进行数据规约, 提升数据密度, 减少数据的维度。

### 2.3 知识融合模型

结合微波组件 PFMEA 先验知识和工艺质量数据, 采用分类算法挖掘工艺质量数据中可辅助失效分析的知识信息。知识融合模型建模流程如图 5 所示, 建模数据包含工序信息、产品信息、问题信息等, 按输入输出数据类型选用

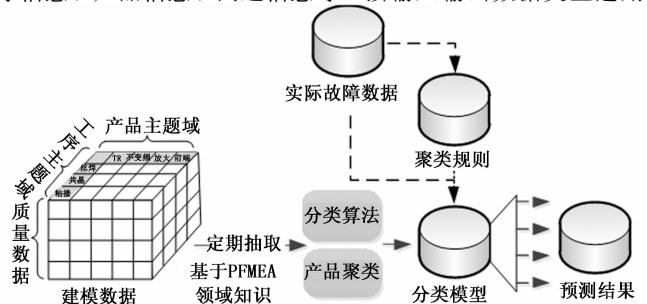


图 5 知识融合模型建模流程

分类算法训练得到分类规则模型或分类黑箱模型, 同时将产品进行聚类后得到的产品类别, 作为分类模型的输入。最后在微波组件工艺问题辅助排故应用时, 通过对输入的产品类别、产品特征信息、故障特征信息、工序特征信息进行运算, 可得到问题分析预测结果。其中, 分类建模算法可采用神经网络、决策树、SVM 支持向量机等模型进行训练<sup>[10]</sup>。

辅助排故模型结构如图 6 所示, 输入为产品编号、产品名称、问题工序、工步、工序特征、失效模式, 输出有故障原因定位、纠正措施、频段、严重度、探测度, 其中加粗的变量名表示向量, 各输入输出数据均为预处理后的标准数据。辅助排故时, 通过输入的产品编号、产品名称, 再从产品/工艺数据库中抽取产品相应工序、工序特征数据对产品进行聚类, 进行相似匹配, 得到产品类别, 再结合输入问题工序、工步、工序特征、失效模式输入到训练好分类模型中, 运算得到按评分排序的问题产品最可能的诱因和处理措施, 实现了对人工处理逻辑和数据信息的融合。

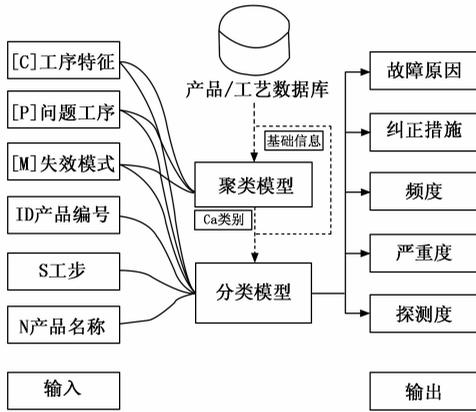


图 6 微波组件辅助排故模型结构

### 3 辅助排故建模实例分析

以某微波组件生产信息系统中的返工返修数据为例,

建立用于对返工返修相关工艺失效问题辅助分析的模型。从数据库中抽取一定时间段内的数据, 样本数据如表 2 所示事务型格式, 主要字段有: 订单号、批次号、产品编号、序列号、产品名称、产品总数、故障代码、故障模式、问题工序、纠正措施、创建时间。

根据 PFMEA 提供的先验知识, 对数据中的各属性字段的潜在关系模式先建立 9 种相关性分析, 从中确定用于聚类的质量特征属性, 如图 7 所示。本例通过逻辑回归对各关键属性进行相关性分析, 得到故障模式与问题原因、纠正措施、工序、问题产品名的相关度 0.62、0.585、0.37、0.1, 纠正措施与工序、故障模式、问题定位的相关度 0.65、0.305、0.22, 问题产品名与工序、问题定位、故障模式的相关度 0.235、0.515、0.175。当取阈值 0.5 时, 即某一因素的相关度大于 0.5, 则认为该因素对分析目标有明显影响, 可作为特征属性, 得出失效模式与工序的相关性较大, 失效模式与问题原因的相关性较大, 纠正措施与工序的相关性较大, 问题产品名与问题定位相关性较大。

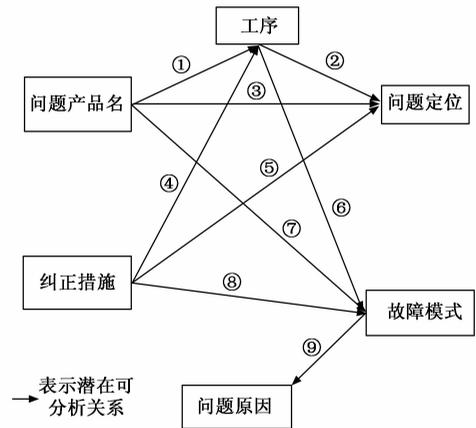


图 7 微波组件质量数据潜在关联关系

表 2 生产系统中抽取经处理的部分样本数据

流水号	产品编号	产品名称	序列号	故障模式	相关工序	问题定位	问题原因	纠正措施
34915	MC3001	开关组件	NA740447	PCB 短路	钳装	驱动芯片	漏锡	更换
34914	MC0022	功放模块	NA186898	器件性能低	共晶	放大芯片	时间长	焊盘去氧化
34913	MC3031	基带模块	NA513683	器件性能低	焊接	低噪放芯片	温度过高	焊盘去氧化
34912	MC3010	下变频组件	NA680065	电路片短路	粘接	电路片	胶溢出	控制胶量
34911	MC1127	开关整件	GX001761	焊点脱落	热声焊	金丝	表面污损	表面清洁
34910	MC3009	开关功分组件	NA789452	电路片脱落	粘接	胶	胶过少	控制胶量
34909	MC3002	校准源模块	GP006221	管芯失效	共晶	FET 管芯	时间长	焊盘去氧化
34908	MC3036	TR 组件	NB142491	器件性能低	封盖	裸管芯	操作损伤	更换
34907	MC3036	TR 组件	NB142510	器件性能低	粘接	开关管芯	方向粘反	加强识图
34906	MC3038	TR 组件	NB142607	组件自激	粘接	内埋滤波器	应力开裂	优化温度曲线
34905	MC1101	均衡模块	GE088207	电路片短路	粘接	胶溢出	胶未修净	控制胶量
34904	MC3157	前端组件	NA442005	污损	微隙焊	芯片电容	元器件损伤	筛选
34903	MC3691	上变频组件	NA115971	性能低效	热声焊	金丝	金丝漏焊	加强识图
34902	MC3691	上变频组件	NA115968	焊点脱落	热声焊	金丝	镀层厚度超差	优化焊接参数

由于微波组件产品典型的种类多、批量小的特点,问题产品名与工序、故障模式的相关性较低,而在实际排故时多以问题产品名为关键信息,因此,将工序、工序特征、故障模式、故障原因作为聚类特征向量对产品名称进行聚,得到部分聚类结果如图 8 所示。

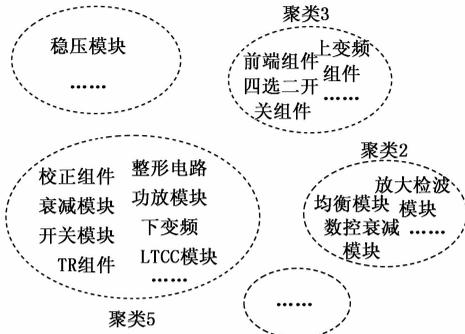


图 8 微波组件相似聚类结果

聚类 2 中的各产品主要特征—微隙焊(工序)、微金丝金带(工序特征)、焊点脱落(故障模式)、镀层厚度超差(故障原因)具有较强相似性;聚类 3 中的各产品主要特征—粘接(工序)、微波印制电路板与铝合金(工序特征)、电路板脱落(故障模式)、胶过少(故障原因)具有较强相似性;聚类 5 中的各产品主要在以下特征上具有较强相似性—热声焊(工序)、镀金电路板与芯片(工序特征)、焊点脱落(故障模式)、镀层不匹配(故障原因)。因此将聚类分析中的主要特征属性作为该类的标签,并作为知识模型聚类输出结果以及分类的输入变量。显然,由于产品与特征属性的高端相关性,具有相同工序、相同工序特征、相同故障模式和原因的产品聚为一类后,可实现数据密度的提升。

前文分析中,由于辅助排故模型每一个输出都有 2 个以上的可选值,所以该分类模型属于多元分类,此处采用所有对所有(AVA)方法<sup>[11]</sup>,即针对输出的某个值为正例,其他值为负例,依次将每个值分别作为正例,其他为负例,对所有输出建立  $m(m-1)/2$  个 2 元分类模型,其中  $m$  表示各输出的取值个数,采用决策树算法实现该模型,并增加纠错码提升准确性。如图 9 所示,以纠正措施中的“更换”为正例,其余纠正措施定义为“返工”,为负例,对生产系统中抽取的质量数据建立了从产品依次经工序、工序特征、故障模式判断得到纠正措施的 2 元决策树模型。同样可将其他纠正措施作为正例依次建立相应的 2 元决策树模型,或对故障原因建立 2 元决策树模型。以图 9 中第一个决策结点为例,初始数据中有 60% 样本的纠正措施是“更换”,当输入产品类别时,如果产品是“TR”则可以占总样本中的 17% 以 90% 精度划分出“更换”,再通过第二级结点工序不是“焊接”判断出总样本中的 29%,并以 92% 精度划分出“更换”,直到最后一级故障模式,完成样本的分类。

根据混淆矩阵计算评估决策树模型的性能,训练得到的决策树模型整体准确率达到 84.68%,召回率为 89.38%,该模型整体性能较好,但还可以通过结合其他维度信息调

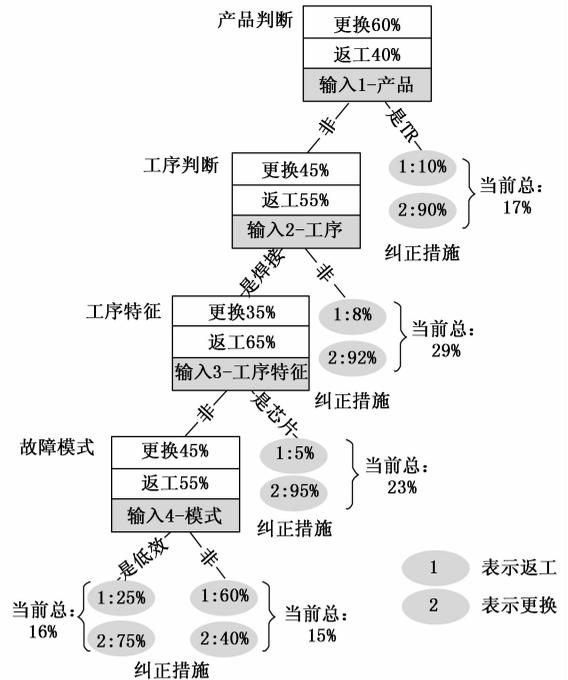


图 9 微波组件辅助排故纠正措施决策书模型

整决策树节点的代价敏感系数,进一步降低模型对“返工”的误判率。

表 3 决策树模型混淆矩阵

		预测类	
		类=2	类=1
实际类	类=2	75.08%	8.92%
	类=1	6.4%	9.6%

最后,通过开发一套质量数据分析软件对上述模型部署应用,如图 10 所示,在软件的辅助排故流程的向导式录入界面,微波工艺人员输入产品及故障相关基本信息数据,软件将自动利用由历史故障数据训练好的知识融合模型进行计算,进而推送给出失效原因、改进措施等信息,方便实现对产品生产中出现问题快速排查和处理。



图 10 基于质量数据模型的辅助排故软件