

基于 CEM-IG 算法的复杂产品关键质量特性识别

闫伟, 何桢, 李岸达

(天津大学 管理与经济学部, 天津 300072)

摘要 为了提高复杂产品高维不平衡质量特性数据集关键质量特征识别效率, 提出 CEM-IG 识别方法. 通过调整 CEM (classification EM algorithm) 算法的 K 值输出不同的聚类结果, 消除冗余样本后作为 IG (information gain) 算法的输入, 并以 IG 作为判别质量特性重要程度的标准构建识别模型, 最终输出最优关键质量特性集. 算例结果表明, 该方法将 CEM 的缺失值处理能力和 IG 的不相关特性筛选能力优势互补, 能够有效降低不平衡和高维度带来的负面影响, 正确识别产品关键质量特性.

关键词 复杂产品; CEM-IG; 高维; 不平衡数据集

Identification of critical-to-quality characteristics for complex products using CEM-IG algorithm

YAN Wei, HE Zhen, LI An-da

(College of Management and Economics, Tianjin University, Tianjin 300072, China)

Abstract In order to improve the efficiency of critical-to-characteristics identification in high-dimensional imbalance data sets for complex product, the CEM algorithm is integrated with IG algorithm, that is, adjusting the K -values in CEM algorithm to get the different clustering results which are inputted for after step, obtaining the recognition model based on CTQ features optional set formed by the IG as to the standard for the discriminate importance of quality characteristics, and selecting the optimal CTQ sets after tests. The results showed that, this method combined the missing data processing power of CEM and irrelevant characteristics screening power of IG successfully, and can reduce the negative effects effectively from imbalances, and can high-dimension and correctly identify the CTQ.

Keywords complex products; CEM-IG; high-dimensional; imbalance data sets

1 引言

在复杂产品生产制造中, 由于复杂产品具有组织结构复杂的特点^[1], 从产品构成的角度来看, 在产品被从整体分解到部件并从部件分解到零件的过程中, 产品质量特性数据集的维度会随着产品结构的逐步分解而不断升高^[2]. 例如, 涡喷发动机从组成结构上被分解为进气装置, 离心压气机, 分管燃烧室, 单级燃气涡轮以及尾喷管等部件, 每个部件都有相应的对其质量进行描述的部件级质量特性 (例如尾喷管的精铸尺寸等等). 同时, 部件又会被进一步分解为轴承, 叶片, 装配螺丝, 导热铜线等零件, 每个零件都有相应的质量特性 (例如拟合度, 重量等等). 层级结构从上到下的分解使得复杂产品的零件级质量特性具有较高的维度, 而高维度的数据集会使产品的质量成本几何级数增长^[3]. 为了降低成本, 就要从产品原始高维度质量特性数据集中识别出对产品质量具有重要影响的少数关键质量特性 (critical-to-quality characteristics, CTQ), 实现在对产品质量的有效控制的基础上降低控制成本.

2 相关研究及问题描述

在复杂产品 CTQ 识别过程中, 许多其他领域的方法被相继引入 (如 ReliefF^[4] 和 Information Gain^[5-6])

收稿日期: 2011-10-27

资助项目: 国家杰出青年科学基金 (71225006); 国家自然科学基金重点项目 (70931004)

作者简介: 闫伟 (1982-), 男, 河北承德人, 博士, 研究方向: 质量工程, E-mail: cdxyw@yahoo.com.cn.

等). 然而, 当这些方法应用在真实生产中时却面临着一些实际困难, 其中之一就是产品原始数据的不平衡性. 在制造业中, 若生产出的合格产品的数量是不合格产品数量的十倍以上, 就称这样的数据为不平衡数据 (im-balance data sets, IDS)^[7]. 不平衡数据 CTQ 识别就是指在各类样本数量差别很大的情况下, 有效识别出对于质量控制具有决定性影响的关键质量特性的过程.

基于 ReliefF 的 CTQ 识别算法是一种处理复杂产品高维平衡数据集的有效方法^[4]. 这种方法, 通过计算产品质量特性与合格产品和不合格产品两个类别内最近样本对应质量特性之间的假设间距来给予每个质量特性相应的权重, 并以此判断质量特性与所属类别之间的相关程度^[8]. 但是当这种方法应用于不平衡数据集时, 因为其中一类产品数量相对太少, 使得在通过计算假设间距来给予相应质量特性权重的过程中, 没有考虑原始样本分布的不平衡性带来的负面影响, 而依然以在识别正确率为判定识别模型效率的标准, 导致模型的建立过分依赖样本数量多的类别, 忽略了样本数量少的类别, 使得识别效果大幅下降. 例如在一个由 90 个合格产品样本和 10 个不合格产品样本组成的不平衡数据集中, 如果以识别精度为判别标准的话, 识别算法会选择将所有的产品样本全部识别为合格产品, 以此来获得 90% 的高准确率. 但显而易见, 这样的识别对于指导实际生产中的质量控制没有是任何意义的.

对于复杂产品生产企业, 少数不合格产品相比于多数合格产品, 通常具有更加重要的意义, 即复杂产品生产企业宁愿把合格产品误识别为不合格产品, 也不希望把不合格产品误识别为合格产品. 因为实际装备复杂产品使用中, 后者带来的危险是前者无法比拟的. 例如, 企业误把不合格的发动机作为合格发动机组装到机身上, 这对于单引擎发动机飞机来说具有致命的潜在威胁, 甚至有可能造成无法估量也无法挽回的损失, 从而导致企业质量信誉下降, 成本急剧增加. 所以, 一方面为了有效识别产品 CTQ, 另一方面为了保证少数类产品的原始数据特性, 将 CEM 算法^[9]与 IG 算法进行整合并重新构建 CTQ 识别算法框架, 使两种方法优势互补, 提出了一种适用于高维度不平衡数据集的复杂产品 CTQ 识别方法. 算例结果表明, 该方法能够有效克服传统方法在处理数据不平衡方面的局限性, 达到提高识别效率, 改善识别效果的目的.

3 构建基于 CEM-IG 算法的 CTQ 识别方法

本部分从与传统方法不同的视角, 构建了一套完整的高维不平衡数据 CTQ 识别方法: 从样本数据集内部类别的视角尝试去发现占有绝大多数比例的数据类中样本所表现出来的差异性和近似性, 将相似性很高的样本聚类, 以此消除冗余样本. 然后将处理后的数据输入信息增益算法, 并建立判别标准. 通过对限定条件下所有可能情况进行迭代计算, 找出最优 CTQ 数据集. 方法的基本构架如图 1 所示.

基于 CEM-IG 的 CTQ 识别方法步骤如下:

- Step 1 收集产品原始样本质量特性数据集.
- Step 2 初始质量特性筛选.
- Step 3 将合格产品数据集与不合格产品数据集分开.
- Step 4 使用 CEM 方法对合格产品数据集聚类.
- Step 5 消除冗余样本.
- Step 6 计算每个质量特性的 IG, 并将结果按 IG 顺序排列.
- Step 7 CTQ 备选数据, 构建识别模型.
- Step 8 测试 CTQ 备选数据集识别效果, 输出 CTQ 质量特性数据集.

该方法的构建从大的功能角度划分为三个阶段. 第一个阶段 (Step 1~3) 对原始高维质量特性数据集进行初始筛选; 第二阶段 (Step 4~6) 对样本不平衡数据集进行 CTQ 质量特性识别; 第三阶段 (Step 7~8) 验证 CTQ 的有效性, 并输出 CTQ 数据集.

3.1 第一阶段: 数据初始处理

这个部分对原始样本的高维度质量特性数据集进行初步维处理, 包括方法构架中的 Step 1~3.

Step 1 收集产品原始样本质量特性数据集. 在产品质量特性收集阶段, 要遵循只有描述产品的特性集包含了足够的描述信息, 才能够准确识别 CTQ 的原则. 因为从理论上讲, 如果两个样本的所有质量特性参数都相同的话, 那么产品质量也应该是相同的. 所以, 为了使原始样本数据集中包含足够的信息, 就需要在本阶段尽可能多的收入质量特性.

Step 2 初始质量特性筛选. 在复杂产品制造生产中收集到的产品质量特性数据集, 因为各种不确定因

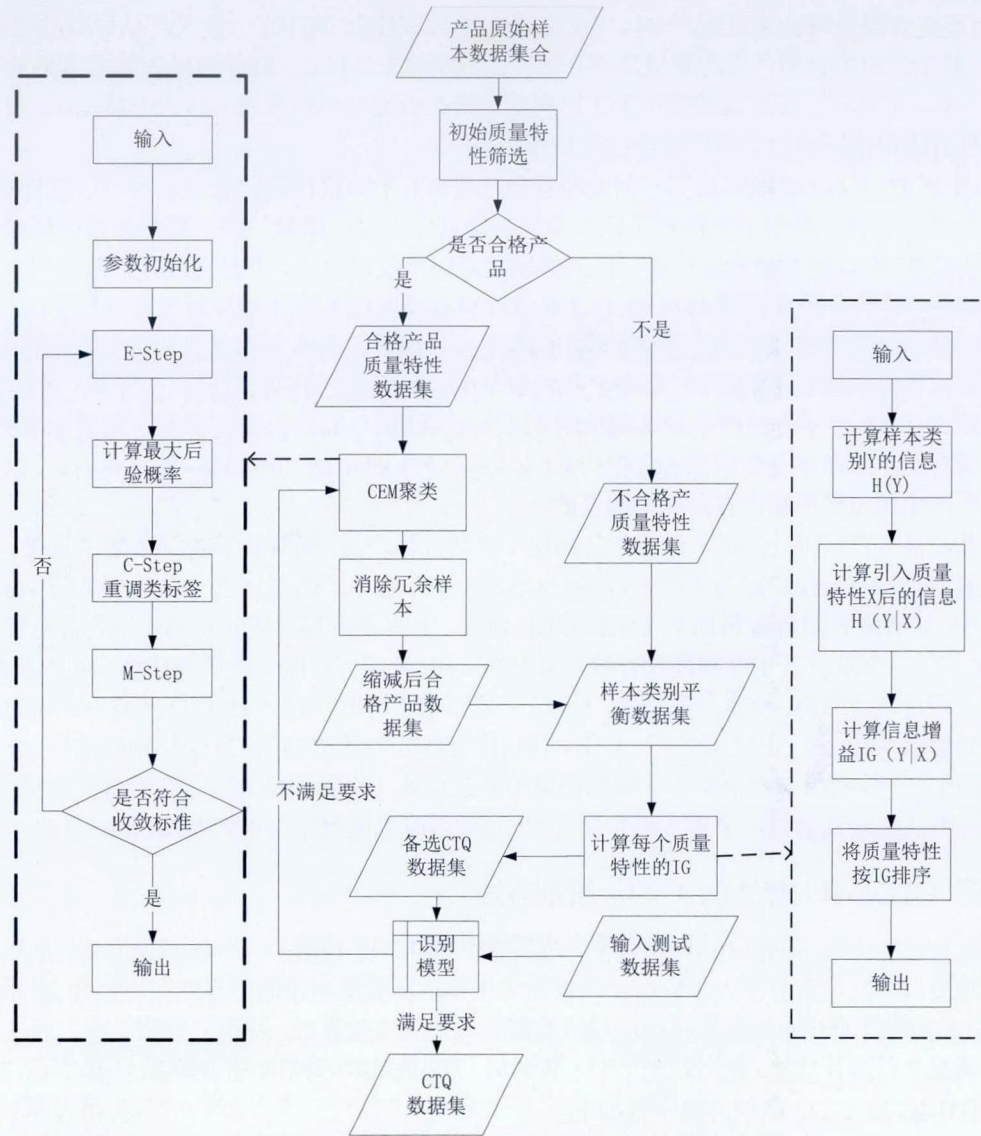


图 1 CEM-IG 关键质量特性识别

素的存在, 往往包含了不完整数据, 噪音数据, 以及其他类型的缺失数据. 在质量特性识别中, 如果一个特性的缺失值过大, 就会对识别规则的建立带来负面作用, 所以要先处理原始数据中的这些缺失值. 计算每个质量特性的缺失值比例 $a\%$, 并根据样本数据集的特点确定阈值.

$$\text{如果} \begin{cases} a\% > \tau, \text{则删除该质量特性} \\ a\% < \tau, \text{则保留该质量特性} \end{cases} \quad (1)$$

Step 3 将合格产品数据集和不合格产品数据集区分开, 并记录每个类别内产品的数量.

3.2 第二阶段: 不平衡数据 CTQ 识别

本阶段主要包括两个部分, 第一, 对不平衡数据进行平衡化处理, 第二, 对所有质量特性依据其与所属类别的相关程度大小依次排序. 包括方法构架中的 Step 4~6.

Step 4 对合格产品聚类. 在第一阶段处理后的样本数据集中, 虽然缺失值比例较大的质量特性已经被删除, 但是仍然有很多质量特性具有一定的缺失比例, 而一般的聚类算法是不能处理包含缺失值数据的. 为了解决这个问题, 本步采用了可以有效处理具有缺失值数据的 CEM 算法进行聚类. 假设 R 为包含 n 个样本的数据集合, 由样本 X_1, X_2, \dots, X_n 构成. k 表示全部 $K(K \in \{1, 2, \dots, k, \dots, NCNumber\})$ 个聚簇中的第 k 类, 表示各个类别所占比例的权重 $0 < p_k < 1 (k = 1, 2, \dots, K)$ 且 $\sum_k p_k = 1$. $f(x, a_k)$ 表示每个聚类分布的密度函数, 其中 p_k^m 和 a_k^m 为未知变量, a_k^m 为第 k 个类别内的分布参数. 计算过程如下:

输入: 原始样本数据集.

输出: 带有类别标签的样本数据集.

1) 初始化: p_k^m, a_k^m .

2) Repeat.

E-Step: 从第一次估计 p_k^m 和 a_k^m 开始, 计算后验概率 $t_k^m(x_i)$ ($i = 1, 2, \dots, n, k = 1, 2, \dots, K$) 即将 x_j 划分为第 k 个聚类 p_k 的概率.

$$t_k^m(x_i) = \frac{p_k^m f(x_i, a_k^m)}{\sum_{k=1}^K p_k^m f(x_i, a_k^m)} \tag{2}$$

C-Step: 将样本 X_j 划分到 $t_k^m(x_i)$ 最大的聚类 p_k 中, 得到样本类标签计为 P_k^m .

M-Step: 对每个样本的 P_k^m 进行评估, 对 $k = 1, 2, \dots, K$, 计算最大似然估计 (p_k^{m+1}, a_k^{m+1}) 判断其是否收敛.

$$P_k^{m+1} = \frac{\#P_k^m}{n}, \quad k = 1, 2, \dots, K \tag{3}$$

$\#P_k$ 表示第 k 个聚类内的样本数量.

Until 满足收敛标准.

3) 输出每个样本的类别标签.

Step 5 消除冗余样本. 根据不合格产品样本的数量确定合格产品平衡化处理目标样本数量 $CN = [CN(1) + CN(2) + \dots + CN(i) + \dots + CN(K)]$. 通常在两个类别数量接近的时候, 识别效率相对较高, 识别效果相对较好. 所以, 在每个类别中取前 $CN(i)$ 个样本. $SampleNumber$ 为全部合格产品的数量, $NCNumber$ 为不合格产品数量.

$$CN(i) = \frac{Cluster(i)}{SampleNumber} \times NCNumber \tag{4}$$

将合格产品与不合格产品数据集合并.

Step 6 计算每个质量特性的 IG [6]. 假设在 Step 5 处理后的样本质量特性数据集 FS 中, 每个样本具有 j 个维度的质量特性, X_{ij} 表示第 i 个样本的第 j 个质量特性的测量值. 每个产品样本都带有一个类标签 Y_i . 产品样本最终属于哪个类别会受到产品样本质量特性值 $(X_{11}, X_{12}, \dots, X_{1j}, \dots, X_{1m})$ 直接影响. 高维关键质量特性识别是指在已知每个样本的质量特性值 X_{ij} 的情况下, 计算他们属于类别 Y 的条件概率 $P(Y/X)$, 以此来表示样本的不同质量特性与样本质量类别之间的相关程度. 通过在原始数据集中重复测试不同的质量特性以及不同质量特性组合的变化对产品质量是否产生直接关键影响, 最后在其中选出对于样本所属类预测最准确的一组质量特性作为复杂产品的 CTQ 数据集. 计算步骤如下:

定义: 原始特征集为 FS ; 阈值为 μ : 用来确定质量特性按照权重筛选的界限; 筛选后的 CTQ 特性数据集为 $Fsub$; $Fsub$ 内的 CTQ 质量特性数量为 N ; 限定 $Fsub$ 内 CTQ 特征数目为 P : 预先设定的 CTQ 特性数据集大小.

输入: 原始样本集 FS 和阈值 μ .

输出: 质量特性重要程度.

1) 给定样本集, 初始关键质量特性集 $Fsub = \{ \}$;

2) 在给定质量特性值的情况下计算对于样本质量类别的信息增益 $IG(Y/X)$; 计算 Y 的信息熵:

$$H(Y) = - \sum_j P(y_j) \log_2(P(y_j)), \quad j \in \{1, 2, \dots, n\} \tag{5}$$

将质量特性变量 X 引入类别变量 Y :

$$H(Y/X) = - \sum_j P(y_j) \sum_i P(y_j/x_i) \log_2(P(y_j/x_i)), \quad i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, m\} \tag{6}$$

其中 $P(y_j)$ 代表类别变量 Y 的先验概率, $P(y_j/x_i)$ 代表引入质量特性变量 X 后类别变量 Y 的后验概率. 引入质量特性变量 X 的信息后, 类别变量 Y 的信息熵 $H(Y) \geq H(Y/X)$, 即引入 X 后, Y 的不确定程度会变小或保持不变. 若 X 与 Y 不相关, 则 $H(Y/X) = H(Y)$; 若 X 与 Y 相关, 则 $H(Y/X) < H(Y)$, 而差值 $H(Y) - H(Y/X)$ 越大, Y 与 X 的相关性越强. 因此, 式 (7) 定义信息增益 $IG(Y/X)$ 为 $H(Y)$ 与 $H(Y/X)$ 的差值, 反映了 X 与 Y 的相关程度, $IG(Y/X)$ 越大, 则变量 X 与 Y 的相关性越强.

$$IG(Y/X) = H(Y) - H(Y/X) \tag{7}$$

3) 将全部质量特性按照 IG 值从大到小排序, 以此表示其重要程度.

3.3 第三阶段: 测试识别效果, 输出 CTQ

本阶段对不同质量特性数据集进行验证并判断是否为最优质量特性数据集, 最后输出 CTQ, 包括方法构架中的 Step 7~8.

Step 7 记录 CTQ 备选数据集, 构建识别模型. 根据 IG 的排序, 依次选取前 $P(P \in \{1, 2, \dots, \text{all.feature}\})$ 个质量特性形成 CTQ 备选数据集.

Step 8 测试 CTQ 备选数据集识别效果, 输出 CTQ 数据集. 通过迭代计算 CEM 中不同的 K 值以及 Step 7 中不同 P 值情况下的备选数据集识别效果, 选出最优, 输出相应的 CTQ 数据集. 在对识别效果进行评价时将第二类错误率引入不平衡数据 CTQ 识别判别准则. 第二类错误描述如下:

在代价矩阵 (表 1) 中, 每列代表由分类规则预测的分类, 每行代表真实的类别, 其中: “ TP ” 表示准确预测的合格产品数; “ TN ” 表示准确预测的不合格产品数; “ FN ” 表示预测为不合格, 而实为合格品的数目; “ FP ” 表示预测为合格, 而实为不合格品的数目 [10].

表 1 代价矩阵

		预测分类	
		合格产品	不合格产品
真实分类	合格产品	TP	FN
	不合格产品	FP	TN

在复杂产品 CTQ 识别中, 把实际为合格品判断为不合格的错误称为 “第一类错误”, 用 α 表示; 将把实际为不合格品判断为合格的错误称为 “第二类错误”, 用 β 表示. 由于第二类错误所带来的损失通常远高于第一类错误 (例如企业把不合格陀螺仪作为合格品交付并安装在飞机上, 其带来的后果及损失将远远大于将合格产品识别为不合格产品带来的影响) 所以企业更愿意以可以接受的 α 升高来换取 β 的大幅下降. 两类错误概率如下:

$$\alpha = \frac{FN}{TP + FN}, \quad \beta = \frac{FP}{TN + FP} \tag{8}$$

4 算例应用

数据选自 UCI 数据库 SECOM 数据集 [11]. 该数据集有 1567 个样本, 每个样本有 590 个质量特性, 标号为 “No.1” 至 “No.590”; 样本分为合格产品和不合格产品 2 类, 其中不合格样本数量为 104, 合格样本数量为 1463 个. 显然, 原始数据是具有高维度质量特性的样本类别比例不平衡数据集. 算例分别使用传统方法 ReliefF 方法 [4] 和 CEM-IG 方法对 SECOM 数据集进行 CTQ 识别, 并将结果从降维水平、分类器、分类精度、第二类错误率 [12] 等方面进行比较评价, 分类器采用 Bayes Classifier.

4.1 传统 CTQ 识别方法

本部分应用基于 ReliefF 算法的 CTQ 识别方法直接对原始的高维度不平衡数据进行 CTQ 识别, 识别正确率如图 2 所示, 第二类错误如图 3 所示, 并在计算过程中分别截取了当 CTQ 数据集内质量特性数量为 11, 21, 31, 41, 51, 61, 71, 81, 91, 101 时的识别结果列入表 2.

表 2 原始数据 CTQ 识别结果

Dataset N	11	21	31	41	51	61	71	81	91	101
CTQ	No.52 等	No.128 等	No.175 等	No.189 等	No.15 等	No.140 等	No.9 等	No.540 等	No.10 等	No.393 等
CCR(%)	93.4	93.4	93.4	93.4	93.4	93.4	93.4	93.4	93.4	93.4
TYPE II Err	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

注: CCR 为识别正确率. TYPE II Err 为第二类错误.

4.2 基于 CEM-IG 的 CTQ 识别方法

本部分用 CEM-IG 方法对原始高维不平衡数据进行 CTQ 识别. 因为不合格产品数量 $NCNumber = 104$, 所以令算法中阈值 $\tau = 70\%$, $K \in \{1, 2, \dots, k, \dots, 104\}$. 识别正确率如图 2 中的所示, 第二类错误如图 3 所示, 并在计算过程中分别截取了当 CTQ 数量为 11, 21, 31, 41, 51, 61, 71, 81, 91, 101 时的识别结果列入表 3, 其中识别正确率最高点出现在 $DataSet.N = 71$ 时, 识别正确率最高为 76.1%, 第二类错误率为 25%, 此时输出包含 71 个关键质量特性的最优 CTQ 数据集: (No.175, No.65, ..., No.113, No.578, ..., No.324,

No.103).

表 3 平衡数据 CTQ 识别结果

Dataset N	11	21	31	41	51	61	71	81	91	101
CTQ	No.34 等	No.35 等	No.56 等	No.446 等	No.173 等	No.455 等	No.175 等	No.37 等	No.308 等	No.444 等
CCR(%)	69.4	72.1	74.8	72.3	73.1	73.6	76.1	75.6	74.2	74.3
TYPE II Err	34%	35%	31%	30%	28%	26%	25%	24%	26%	25%

图 2 和图 3 分别将两种方法在 CTQ 数量依次为 1, 2, ..., all.feature 情况下的分类精度和第二类错误率做了对比, 图中实线表示 ReliefF 方法结果, 虚线表示 CEM-IG 方法结果, 横轴表示 CTQ 数据集内关键质量特性的数量, 图 2 纵轴为识别正确率, 图 3 纵轴为第二类错误率。

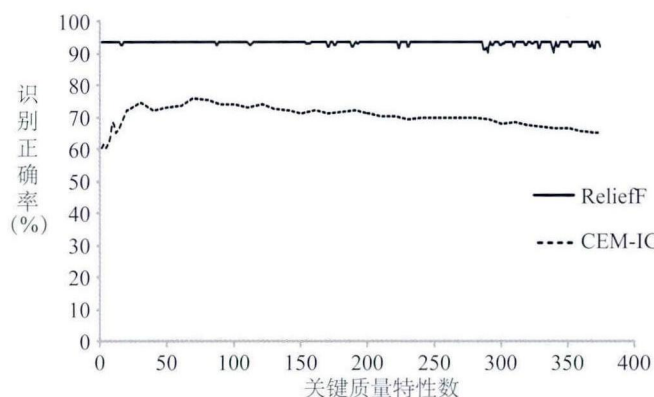


图 2 识别正确率对比

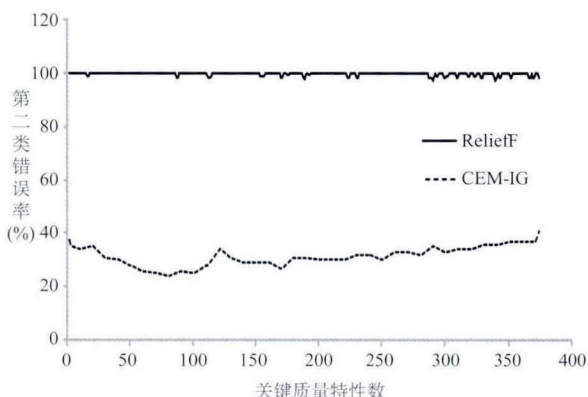


图 3 第二类错误率对比

由以上可以很直观看出来, 传统基于 ReliefF 的 CTQ 识别方法的识别正确率在 93.4% 以下小幅波动, 第二类错误率绝大多数都为 100%, 即使有波动, 波动幅度也极小, 可以近似认为是 100%。

基于 CEM-IG 的 CTQ 识别方法的正确率在 73% 左右波动, 第二类错误率在 30% 左右波动, 最小的时候为 25%。与传统 ReliefF 方法相比, 虽然识别正确率有一定的下降, 但是换来了第二类错误率的大幅降低, 这在实际中是有极其重要的现实意义的, 如同文章前部分所述, 对于复杂产品生产企业来说, 尤其是在定制生产的复杂产品应用中, 错误地将合格产品判定为不合格产品带来的损失要远远小于将不合格产品判定为合格产品的损失。

5 结论

近年来, 随着企业管理水平和制造水平的提高, 产品的缺陷率已经变得越来越低, 合格产品的数量往往占据极大比例, 而不合格产品的数量少之又少, 数据的不平衡性使原本很困难的 CTQ 识别变得更加困难。

为了在抵消数据不平衡带来负面影响的情况下高效地识别出 CTQ 数据集, 从一个与传统关键质量特性识别方法不同的角度提出了复杂产品 CTQ 识别研究路线。算例表明, 该方法可以在大幅降低质量特性维度和大幅降低第二类错误的概率的情况下有效识别产品 CTQ, 是复杂产品高维度不平衡数据 CTQ 识别方法探究的良好范例。

参考文献

- [1] 苏越良, 张卫国. 基于全局的复杂产品开发项目风险协调控制方法 [J]. 系统工程理论与实践, 2008, 28(5): 70-76.
Su Yueliang, Zhang Weiguo. Study on a risk coordination control method based on overall situation for complex product development project[J]. Systems Engineering — Theory & Practice, 2008, 28(5): 70-76.
- [2] 李伯虎. 复杂产品制造信息化的重要技术 —— 复杂产品集成制造系统 [J]. 中国制造业信息化, 2006(14): 19-23.
Li Bohu. Key technologies in informatization of complex product: Complex product integrated manufacturing system[J]. Manufacture Information Engineering of China, 2006(14): 19-23.
- [3] 张霖, 任磊, 陶飞. 复杂产品制造数字化集成平台技术 [J]. 国防制造技术, 2010(4): 5-10.
Zhang Lin, Ren Lei, Tao Fei. Complex product manufacturing digital integrated platform technology[J]. Defense Manufacturing Technology, 2010(4): 5-10.
- [4] Yan W, He Z, Tian W M. The application of ReliefF algorithm for identifying CTQ in complex products[C]//

- 2011 2nd IEEE International Conference on Emergency Management and Management Sciences, Beijing: 2011: 459-463.
- [5] 徐济超, 马义中. 多指标稳健设计质量特性的度量 [J]. 系统工程理论与实践, 1999, 19(8): 45-48.
Xu Jichao, Ma Yizhong. Measurement methods for robust multivariate quality characteristics[J]. Systems Engineering — Theory & Practice, 1999, 19(8): 45-38.
- [6] 闫伟, 何桢, 田文萌, 等. 基于 IG 的复杂产品关键质量特性识别 [J]. 工业工程与管理, 2012, 17(1): 70-74.
Yan Wei, He Zhen, Tian Wenmeng, et al. Complex products critical-to-quality characteristics identification based on IG[J]. Industrial Engineering and Management, 2012, 17(1): 70-74.
- [7] 张建, 方宏彬. 剪枝与欠采样相结合的不平衡数据分类方法 [J]. 计算机应用研究, 2012, 29(3): 847-848.
Zhang Jian, Fang Hongbin. Pruning and undersampling combination of imbalanced data classification method[J]. Application Research of Computers, 2012, 29(3): 847-848.
- [8] Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF[J]. Machine Learning, 2003, 53(1-2): 23-69.
- [9] 高旅端, 陈志. EM 算法的拟 Newton 加速方法 [J]. 系统工程理论与实践, 1998, 18(10): 69-73.
Gao Lüduan, Chen Zhi. A quasi-Newton acceleration method of the EM algorithm[J]. Systems Engineering — Theory & Practice, 1998, 18(10): 69-73.
- [10] 李作春, 周秀梅, 袁鼎荣. 代价敏感学习的过度拟合问题研究 [J]. 广西大学学报: 自然科学版, 2009, 34(6): 809-814.
Li Zuochun, Zhou Xiumei, Yuan Dingrong. A study on over-fitting in cost-sensitive learning[J]. Journal of Guangxi University: Natural Science Edition, 2009, 34(6): 809-814.
- [11] Murphy P, Aha D. UCIML repository secom dataset[EB/OL]. (2008-09-30). <http://archive.ics.uci.edu/ml/machine-learning-databases/secom>.
- [12] Shen H, Wan H. Controlled sequential factorial design for simulation factor screening[J]. European Journal of Operational Research, 2008, 198(2): 511-519.