Gaussian mixture models-based control chart pattern recognition

基于高斯混合模型的控制图模式识别

摘要：控制图中显示的异常模式可以与工艺的变动明确地联系起来。快速、准确的控制图模式识别（CCPR）能够显著地缩小造成异常的原因范围，加快解决问题的处理过程，因此研究CCPR是完全必要的。本文利用多个高斯混合模型（GMM）模型构建CCPR模型，提出了基于GMM的CCPR模型。将统计特征和小波能量特征作为输入变量，使得改善后的CCPR模型较单一CCPR模型训练更加简单和高效，同时模型效果也更加好，因此质量工程师和操作员能够更加方便地使用改进之后的模型。并且，该模型可以通过采用动态计划来很好地适应新的控制图。实验结果表明模型在当前CCPs的检测和识别方面效果良好，并有效地适应新的CCPs。更重要的是，模型为因为其高效计算和良好的识别效果为在线检测提供一种可能。最后本文提供了一个基于GMM统计过程控制（SPC）识别系统的开发方案。

1.简介

统计过程控制（SPC）已经被广泛地应用在监控和控制工业部件生产过程，来达到维持部件的生产率和质量。在实施SPC过程中最核心的部分是控制图的异常变量展示的测量。通常在SPC中，一个过程在只有随机因素影响时会假设其是自然或者正常的。常见考虑的原因是由于其内在的自然或者正常过程。特殊原因被定义为对正常过程的异常冲击，这种冲击需要尽可能快地被标识和清除。当控制图中显示异常就需要确定并清除异常原因。进一步讲，对观察结果进行分析，例如，控制图显示的正常和不正常模式，提供了很多过程信息，该信息可以通过合适的分析来推断出造成这种现象背后的原因。因此，控制图模式（CCPR）识别成了SPC的关键任务。CCPR可以结合工程知识极大地缩小异常原因的范围。这种极大地缩小异常诊断范围的方式给予处于失控状态的过程以必要的修正和调整使之回归受控。然而，控制图再出现异常时不提供任何和原因相关的信息，只根据当前的观测结果监控制造过程。

多年来，像 zone tests ，run rules等很多规则被提出来，来补充对控制图的分析。这些规则来协助质量人员检测控制图上异常模式（CCP）。使用这些规则存在的主要问题有，应用所有规则会造成的过量的假警报（Guh2005）；同时，它往往依赖于技巧和分析的经验来确定是否存在异常的CCP。此外，为了构建新的CCPs需要添加新的规则，这样会使规则集变的很大，从而导致难以维护和管理。因此，CCP在zone tests or run rules 方面有很多困难，因为这涉及到模式识别、自适应学习方面。

由于过程监控中数据量大，手工检查效率低，不一致，现有的基于人工智能的方法为实现过程监控和故障诊断自动化提供了一种有效途径。人工神经网络（ANNs）具有良好的实时噪声容忍度，不需要对监测数据的统计分布进行假设。这一重要特性使人工神经网络成为有前途和有效的工具，可以用来改善制造质量控制应用中的数据分析。在过去的十年中，不同的结构人工神经网络和学习算法在质量控制中得到了广泛的应用。概率神经网络（Plummer 1993）， learning vector quantisation networks (LVQ)，自适应谐振理论（ART）(Hwarng and Chong 1995) ，径向基础函数（RBF）（Cook and Chiu 1998），多层感知器（MLP）神经网络（(Chang and Ho 1999 ）等。已经用来识别基础异常的模式如：移位，趋势，周期，混合。Guh（2004）使用遗传算法来优化CCPR中的ANN误差前馈。Guh（2005）更进一步提出了混合学习模型，整合了BPN和决策树，为CCP时设计了一个高效的识别系统。wang（2007）针对离散过程提出一个包括ANN新的框架能够识别6种过程信号。他们的结果表明，通过小波滤波去噪可以有效地提高用于识别过程模式和ART神经网络识别性能。Jiang（2009）提出了一种神经网络数值拟合模型来识别CCPs。在这个模型中，使用数值拟合方法来估计参数和模式的特定类型，这是不同于常规的人工神经网络的分类方法。近年来，人工神经网络集成方法已被用于制造和诊断监测过程(Yu and Xi 2009a,b, Yu et al. 2009).

ANNs的一个缺点是人工神经网络的过程信息是不可见的，使用者无法理解内部原因。另外，网络结构一般比较复杂，训练过程耗时长。对此，为了解决这些问题提出了基于特征的分类方法(Hassan et al. 2003, Assaleh and Al-Assaf 2005, Gauri and Chakraborty 2007, Yu et al. 2007)，实验结果表明，特征提取的识别器进行显式结构和显著减少了训练时间。

近年来，其他的识别器（如，决策树，(Wang et al. 2008) ；c-means， (Wang and Kuo 2007) ；支持向量机(SVM) (Ranaee et al. 2010)；统计相关系数方法，(Yang and Yang 2005) 等）已经被应用到CCPR。虽然基于这些分类模型的识别器表现出良好的性能，但其结构一般都是固定的一旦建成使用，由于复杂的动态工作条件，时常产生新的异常模式，可能会限制他们在真实世界的使用。

根据以上的分析可以看出，快速准确地识别关键控制点是非常必要的，尤其是在实际的制造过程。此外，新型CCPS在线识别仍然是需要解决的一个重要问题。类比其他相近的研究(Hassan et al. 2003, Guh 2005, Gauri and Chakraborty 2007, etc.), 本研究还考虑了以下八种有意义的模式：正常状态、上下转换、上下趋势、周期、混合和系统模式。

本研究中，在独立离散过程中使用了基于高斯混合分布（GMM）的模型。GMM一个是对估计过程观测很有效果的概率密度函数（PDF）。GMM是基于无监督学习的模型，模型会根据输入数据的性质进行拟合，学习复杂的分布。模型的训练完全是数据驱动的。与之前使用单一的识别器识别所有的CCP不同，针对在制造过程中发生异常，我们构建了6个GMM模型。这样的模型构造的方案在应用到实际制造过程中很有吸引力。事实上，只使用一个CCP的知识来构建一个GMM，因此比所有六个点的训练数据训练的GMM更好的泛化性能。此外，基于GMMs的分类模型是不固定的，可以根据实际应用的要求改变。例如，它可以为新的CCP添加新的高斯分布，改变CCP只需训练一个GMM即可。根据我们的发现，很少有论文论述这种建模方案对提高制造过程适用性的重要性。更重要的是，现有的大多数研究往往只考虑在预先定义（已知）的模式下的识别性能，但没有考虑到自适应能力相对于在制造过程产生的新的（未知）的模式情况。因此，本文的研究目标如下：

（1）基于时间和时频域改进GMM，减少控制图模式识别的建模时间并提高其识别性能。

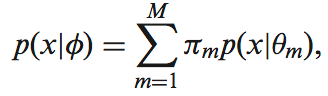
（2）设计一个基于多个GMMs的CCPR模型不仅是结合当前的模型，也是适应未来的新模式；

（3）解决该模型的在线识别性能，包括过程监视和识别。

论文的其余部分安排如下。第2节描述了GMM和相关算法。3部分提出了一种基于GMM模型的新分类。模拟过程用于说明第4节中所提出方法的性能。第5节给出结论并进行延伸讨论。

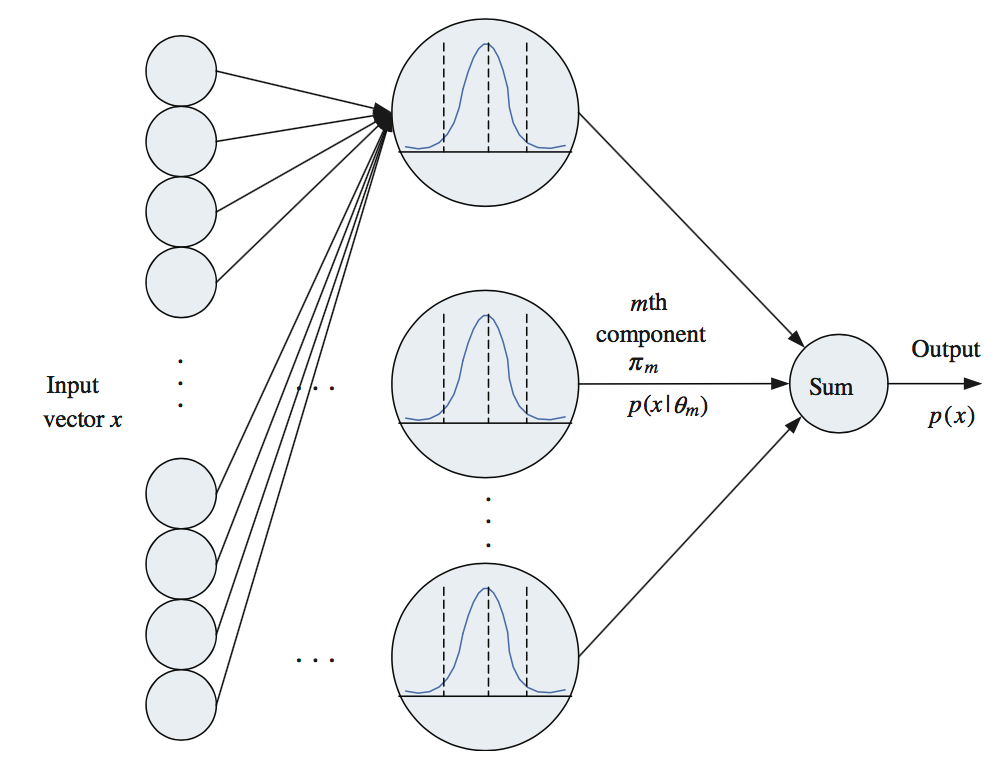
2 高斯混合模型

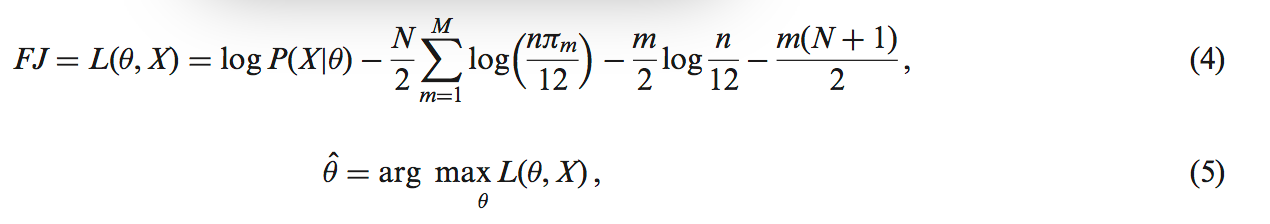
本文，提出了八个关键的控制点（即正常，向上/向下移动，向上/向下的趋势、周期、系统和混合是在3.1节中讨论），基于这些关键控制点建立GMM的识别器。我们先建立6个GMMs分类器，然后识别前6个CCPs。最后新添加的两个关键控制点作为检测的自适应学习能力，提出基于GMM模型的新CCPs国际。有关GMM的详细信息如下所示。

让 是一个d维的随机变量， x 代表一个特定的输出。也即是说x服从一个有限的混合分布，其概率密度函数（pdf）可以写成有限个已知分布的加权的和。比如，每个分布都是高斯分布，x服从高斯混合分布。GMM

差一行

通常的选择是用期望最大化算法（EM）获取优化的混合参数向量。EM是一种强大的统计工具，用于寻找涉及可观测变量和隐藏变量问题的最大似然解。EM是一个迭代的过程，不断的寻找局部最优解。其中E-step和M-step被不断的迭代进行。对于高斯混合模型，EM算法的收敛性得到了很好的研究。(Dempster et al. 1977)

混合学习的标准EM有一个弱点：它需要M（即分布的数量）的知识，以达到一个良好的局部最优。为了克服这一困难，目前已经提出了许多确定标准来估计GMM中适当的分布个数数。大多数实际的模型选择技术都是基于最大化以下类型的准则：

当log 是给定数据集的对数似然，该部分可以使用EM算法来使之最大化。但是，添加更多的分布数肯定是可以增加对数似然，为了达到均衡需要引入惩罚项来惩罚分布的复杂度。模型复杂度该如何选择已经有不少研究了，比如：赤池信息准则（AIC）（赤池1974），最小描述长度（MDL）（rissansen 1987），贝叶斯推理的标准（BIC）（- 1978），在这项研究中，我们使用了Figueiredo和Jain（2002）提出的FJ规范：

其中N是指定每个分布的数量，n是训练样本的总数。Figueiredo和Jain的实验结果表明，该标准比其他标准如MDL，BIC的效果要更好一些。一般来说，为GMM选择一个好的衡量指标是很困难。一个有效的方法是根据先前的经验来从AIC，BIC，FJ中选择。我们之前的实验结果表明，FJ标准的效果优于其他标准（如AIC和BIC），因此作为本研究的模型选择标准。