

分类号: TM391

单位代码: 10335

密 级: 公开

学 号: 21432041

浙江大学

硕 士 学 位 论 文



论文题目 基于 KECA 的非线性故障检测

作者姓名 秦家祥

指导教师 杨春节 教授

学科 (专业) 控制科学与工程学

所在学院 控制科学与工程学院

提交日期 二〇一六年十二月二十六日

A Dissertation Submitted to Zhejiang
University for the Degree of
Master of Engineering



TITLE: Nonlinear Fault Detection Based on KECA

Author: Jiaxiang Qin

Supervisor: Prof. Chunjie Yang Supervisor

Subject: Control Science and Engineering

College: College of Control Science and Engineering

Submitted Date: 2016-12-26

基于 KECA 的非线性故障检测



论文作者签名: _____

指导教师签名: _____

论文评阅人 1: _____ 匿名评审

评阅人 2: _____ 匿名评审

评阅人 3: _____ 匿名评审

评阅人 4: _____

评阅人 5: _____

答辩委员会主席: _____

委员 1: _____

委员 2: _____

委员 3: _____

委员 4: _____

委员 5: _____

答辩日期: _____ 2017-03-10

Nonlinear Fault Detection Based on KECA



Author's signature: _____

Supervisor's signature: _____

Thesis Reviewers 1: _____ anonymous reviewer

Reviewers 2: _____ anonymous reviewer

Reviewers 3: _____ anonymous reviewer

Reviewers 4: _____

Reviewers 5: _____

Committee Chairperson: _____

Committeeman 1: _____

Committeeman 2: _____

Committeeman 3: _____

Committeeman 4: _____

Committeeman 5: _____

Date of oral defence: _____ 2017-03-10

目录

第 1 章 基于 MVU_KECA 的非线性故障检测 1

1.1 引言 1

1.2 最大方差展开 2

1.3 KECA_MVU 4

第 1 章 基于 MVU_KECA 的非线性故障检测

摘要 针对工业过程中的非线性检测问题，一种常用的解决方法是引入核函数，利用核函数实现对原始数据进行高维映射，从而将非线性问题近似线性化。然而，到目前为止，关于核函数中参数选择问题没有统一的结论。本章将最大方差展开 (MVU) 引入到核熵成分分析 (KECA)，利用 MVU 学习得到核矩阵，提出了一种基于 KECA_MVU 的非线性故障检测方法。其主要思想是在利用 KECA 实现检测过程中，其核矩阵由原来的直接通过核函数计算替换为通过 MVU 解优化得到，再通过线性回归近似得到原始数据到降维输出的非线性映射，最后利用该映射实现在线检测，从而避免了核参数选择的盲目性。此外，该方法也继承了 MVU 的特性，能够在降维后实现数据分布边界的保持。在 TE 过程的仿真结果也验证了该方法能够有效地实现对故障的检测。

关键词 故障检测；非线性；MVU；线性回归；KECA

1.1 引言

现代工业工程中引入了集散控制系统 (DCS)，利用 DCS 可以有效地对工业过程实现集中管理和分散控制。DCS 大大的方便了数据的采集，然而，采集到的数据由于工业过程中约束条件以及平衡的存在使得数据具有高相关性、高维度、非高斯性等特点。而实际的工业过程中的数据波动可以通过少量的变量就能够进行有效的解释，这就使得这一类通过将高维数据映射到低维而实现提取其主成分的数据处理方法变得流行起来。通过降维将数据分为互补的两个部分：系统部分和残差部分，这两部分分别可以用来反映降维模型内部的波动和模型外部的波动，从而实现了对生产过程的监测。核方法是一种被广泛用来处理非线性降维的方法，该方法有效地通过核矩阵的空间映射将原始数据的非线性分布近似线性化。KECA 就是一种从信息熵角度出发的利用核函数实现非线性数据处理的方法，其具有较好的监测效果。但是，KECA 并没有考虑到数据的边界分布特性，在实际的生产数据中，能够表征其特性的过程数据实际是一个低维结构，而其原始数据结构则是高维度的结构，也就是说反映数据波动特性的成分是嵌入高维空间的低维结构。使用 KECA 对数据进行降维，并不能很好地从数据结构的角度来解释边界分布这个问题。此外，如何确定 KECA 中核函数的最优和参数也没有固定的结论。流行学习是近年来广泛被研究的一类有效地处理非线性数据降维的方法，它们通过在局部和全局分别进行线性和非线性的数据分布假设，揭示了非线性数据的内在规律，然后引入优化求解得到能够保持数据最优分布特

征的低维嵌入表示。最大方差展开 (MVU) 是比较有代表性的一种流行学习方法, 其假设条件是数据在局部满足线性条件, 并且在降维前后局部范围 k -邻近点之间的欧氏距离保持不变。因此, 其目标可以转换为最大化低维数据的邻近点的欧式距离, 这一点与 PCA 的目标一致, 即其目标是通过保持低维数据方差最大实现降维, 再通过引入核函数, 该方法将训练数据的核矩阵求解问题转换为求解一个半定规划问题, 该问题的最优解即为所求核矩阵 K 。MVU 基于几何分布的流形假设可以使得该方法很好地保存数据分布的边界, 而故障检测离线建模阶段就是通过离线数据建立模型从而确定生产过程的波动边界, 这说明 MVU 可以很好地用于过程监测。然而, MVU 通过半定规划得到核矩阵并没有显示的数据映射表达式, 故不能直接应用于过程监测。邵纪东和姚远等人分别提出了使用线性回归和搞死过程模型来得到投影矩阵, 从而解决了 MVU 在用于过程监测时候实现在线监测的问题。

本章提出了一种基于核熵成分分析和最大方差展开的新的降维方法 (KECA_MVU), 并将其用于故障检测。该方法在离线建模阶段利用 MVU 学习得到核矩阵, 再通过选择熵最大的方向将数据降维得到其得分向量, 然后利用线性回归得到原始数据到降维表示的最优线性近似投影, 最后利用该投影实现在线检测。

1.2 最大方差展开

MVU 与 PCA 有着相同的优化目标, 是 PCA 的非线性扩展。其优化目标是使降维空间中相邻点之间的欧式距离最大, 该目标可以转化为使得映射后的方差最大从而实现数据降维, 它本质上是一种特殊的 Kernel PCA。在使用 KPCA 降维过程中, 通过寻找特定的核函数将数据数据投影到高维的特征空间当中去, 而在数据投影过程中可以计算得到核矩阵 K 。MVU 则是通过直接利用训练数据定义核矩阵实现了将输入数据的流形结构展开到核特征空间过程当中的目标, 定义过程可以转换为一个半正定规划 (SDP) 问题, 对 SDP 求最优解即得到了核矩阵 K 。由于 SDP 求解是一个凸优化问题, 从而确保了所获得的核矩阵是全局最优。设训练数据为 $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\mathbf{x}_n \in \mathbf{R}^D$, 将其映射到核特征空间 \mathcal{F} 当中, 映射后的数据表示为 $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n)$, 设 $\mathbf{y}_1, \dots, \mathbf{y}_n$, $\mathbf{y}_n \in \mathbf{R}^d$, 是其降维后的嵌入表示。其中, N 代表样本数量, D 代表输入数据的变量数, d 代表降维后低维嵌入的维数, 且满足 $d < D$ 。MVU 通过映射 Φ 既实现了展开输入数据的流形结构, 又保证了输入数据与输出的低维嵌入数据是局部等距的, 设其局部临近点为 k 。因此, 为了展开流形结构, MVU 的优化目标就可以表示为使得输出的低维嵌入结构中点的欧式距离最大:

$$\Gamma = \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N \|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|^2 \quad (1.1)$$

为了使得式 (1.1) 所表示的值最大, 需要满足以下几条约束条件:

I 局部等距性:

该约束是用来保证核空间当中的局部流形结构。设 $\mathbf{W} \in \mathbf{R}^{N \times N}$ 是二值邻接矩阵, 可以用其值来判断 \mathbf{x}_i 和 \mathbf{x}_j 是否在 k 近邻区域内相邻。因此, 如果 \mathbf{x}_i 和 \mathbf{x}_j 相邻, 有 $\mathbf{W}_{ij} = 1$, 或者 \mathbf{x}_i 与 \mathbf{x}_j 与另一个点的同时相邻, 有 $[\mathbf{W}^T \mathbf{W}]_{ij} = 1$ 。局部等距约束条件就可以表示为:

$$\|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad (1.2)$$

II 中心化:

中心化约束条件的作用是用来消除数据在映射过程中由于旋转和平移等操作所产生的自由度, 可以将其表示为:

$$\sum_{i=1}^N \Phi(\mathbf{x}_i) = 0 \quad (1.3)$$

从公式 (1.1) (1.3) 可知, 式中存在利用等式约束条件对欧式距离的二次方求解最优值, 该过程并不能保证所得结果是全局最优。为了保证所得最优解为全局最优, 对输出低维嵌入空间定义内积矩阵, $\mathbf{K}_{ij} = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$, 公式 (1.2) 的等距性约束以可以由展开为:

$$\|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \mathbf{K}_{ii} + \mathbf{K}_{jj} - 2\mathbf{K}_{ij} \quad (1.4)$$

将内积带入到目标函数 1.1 和中心化约束 1.3, 优化目标可以表示为:

$$\Gamma = \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N \|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|^2 = \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{K}_{ii} + \mathbf{K}_{jj} - 2\mathbf{K}_{ij}) \quad (1.5)$$

同理, 公式 1.3 可以利用内积的形式展开为:

$$\left\| \sum_{i=1}^N \Phi(\mathbf{x}_i) \right\|^2 = \sum_{i=1}^N \sum_{j=1}^N \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) = 0 \quad (1.6)$$

再将公式 (1.6) 带回到公式 (1.5) 中, 优化目标可以表示为:

$$\Gamma = \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{K}_{ii} + \mathbf{K}_{jj} - 2\mathbf{K}_{ij}) = \text{Tr}(\mathbf{K}) \quad (1.7)$$

此时, 最优解问题转化为求解核矩阵 \mathbf{K} 的迹, 为了保证该过程为一个凸优化过程, 需要在上面等式约束的基础上附加一不等式约束条件 $\mathbf{K} \succeq 0$, 也就是要求核矩阵满足半正定

条件。这样，求解 MVU 最优解的过程就变成了一个半正定规划（SDP）问题，其具体形式为：

$$\begin{aligned} \max \quad & Tr(\mathbf{K}) \quad s.t. \quad \mathbf{K} \succeq 0 \\ & \sum_{ij} \mathbf{K}_{ij} = 0 \\ & \mathbf{K}_{ii} + \mathbf{K}_{jj} - 2\mathbf{K}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 \end{aligned} \quad (1.8)$$

其中，对于所有 (i, j) 有 $\mathbf{W}_{ij} = 1$ 或者 $[\mathbf{W}^T \mathbf{W}]_{ij} = 1$ 。以上凸优化有多种解法，本文中对该凸优化求解可以利用工具包 CSDP 实现，其具体介绍可以参考相应的文章。

通过公式（1.8）可以得到 MVU 的最优解 \mathbf{K} ，再对内积核矩阵 \mathbf{K} 进行特征分解，可以得到输出 \mathbf{y}_i 的表达形式。设 $V_{\alpha i}$ 表示特征值 λ_{α} 对应的第 α 个特征向量的第 i 个元素，则核矩阵可以表示为：

$$\mathbf{K}_{ij} = \sum_{\alpha=1}^d \lambda_{\alpha} V_{\alpha i} V_{\alpha j} \quad (1.9)$$

公式中的 d 表示低维嵌入的维数，该嵌入结构与输入 \mathbf{x}_i 是局部 k 邻接等距的，它可以通过输出的第 α 个元素表示：

$$\mathbf{Y}_{\alpha i} = \sqrt{\lambda_{\alpha} V_{\alpha i}} \quad (1.10)$$

由此，该局部等距的低维嵌入结构就通过谱分解被表示出来了。

1.3 KECA_MVU

从上一节的分析可知，MVU 的目标函数依旧是获得方差最大的方向，它是 PCA 的一种非线性扩展。MVU 通过求解 SDP 问题得到全局最优解 \mathbf{K} ，再利用谱分解由特征值和特征向量组合得到了降维后的低维嵌入的表示。本章则是信息熵的角度出发，将利用 SDP 所获得的核矩阵 \mathbf{K} 不再按照特征值的大小进行排序，而是根据瑞利熵的大小进行排序获取其投影方向，然后对非线性数据实现降维。结合上一章所介绍的知识，在特征空间当中利用 MVU 进行数据投影时，将 Φ 投影到第 i 个主成分 \mathbf{u}_i 可以表示为：

$$P_{\mathbf{u}_i} \Phi = \sqrt{\lambda_i} \mathbf{e}_i^T, \quad (1.11)$$

其中， λ_i 和 \mathbf{e}_i 分别表示特征值以及与之对应的的特征向量。而其瑞利熵的估计可以表示为：

$$\hat{V}(p) = \frac{1}{N^2} \sum_{i=1}^N (\sqrt{\lambda_i} \mathbf{e}_i^T \mathbf{1})^2. \quad (1.12)$$

这就不难发现瑞利熵的估计是由所有到 MVU 主成分方向上的投影组成的。而只有当对应主成分轴的 $\lambda_i \neq 0$ 并且 $\mathbf{e}_i^T \mathbf{1} \neq 0$ 才能够使得其对瑞利熵的估计值有贡献。而那些对瑞利熵贡献较大的主成分方向包含了大部分关于由输入数据提供的概率密度函数的形状信息，如果仅仅是通过选择最大的前 d 个特征值，那么从信息熵的角度来看对数据进行投影时候其投影方向可能有并不包含信息的特征向量。

设 U_d 是 KECA_MVU 的数据投影方向，由前面的分析可知， U_d 是由 MVU 投影当中对瑞利熵贡献最大的前 d 个特征向量张成的子空间。为了使得数据投影后能够保存尽可能多的输入数据的信息，则 KECA_MVU 的优化目标可以表示为：

$$\min_{\lambda_1, e_1, \dots, \lambda_N, e_N} \hat{V}(p) - \hat{V}_d(p) \quad (1.13)$$

其解可以表示为：

$$\Phi_{eca} = P_{U_d} \Phi = \Lambda_d^{\frac{1}{2}} \mathbf{E}_d^T, \quad (1.14)$$

其中， Λ_d 和 \mathbf{E}_d 分别表示 MVU 求解得到的核矩阵特征分解之后对应的特征值和特征向量矩阵。而与 Φ_{eca} 对应的瑞利熵的估计可以表示为：

$$V_d(p) = \frac{1}{N^2} \mathbf{1}^T \mathbf{E}_d \Lambda_d \mathbf{E}_d^T \mathbf{1} = \frac{1}{N^2} \mathbf{1}^T \mathbf{K}_{eca} \mathbf{1} \quad (1.15)$$

将公式 (1.15) 带入到公式 (1.13) 当中，KECA_MVU 的优化目标可以表示为：

$$\min_{\lambda_1, e_1, \dots, \lambda_N, e_N} \frac{1}{N^2} \mathbf{1}^T (\mathbf{K} - \mathbf{K}_{eca}) \mathbf{1} = \frac{1}{N^2} \sum_{j=d+1}^N \psi_j \quad (1.16)$$

公式当中的 ψ_j 则与公式 (1.12) 中的 j 最大项相关，也就是通过选择前 d 项最大的瑞利熵的值可以实现数据投影后保留的信息最多，KECA_MVU 通过选择瑞利熵最大的前 d 项作为数据投影的方向实现了数据的有效降维。