

一种新的灰关联分析算法 在软测量中的应用¹⁾

李勇 邵诚

(大连理工大学先进控制研究所 大连 116024)
(E-mail: Leeyon76@sohu.com)

摘 要 软测量是一种有效的在线测量技术,它在工业生产中受到越来越广泛的重视.作为实施软测量的前提,辅助变量的选择至关重要,它将直接影响软测量的效率.灰关联分析是一种研究变量间相关性的新方法,它适于求解不确定性系统中小样本数据之间的关联程度.深入研究了软测量辅助变量选择过程中存在的问题,基于灰关联分析,提出了一致关联度算法,对辅助变量进行选择.并给出了一种实用的冗余变量剔除方法.同时,结合仿真实例对其有效性进行验证.仿真结果表明该方法是有有效可行的,有助于提高软测量的精度.

关键词 软测量,灰关联分析,一致关联度,径向基网络
中图分类号 TP274

Application of a Novel Grey Relation Analysis Algorithm to Soft Sensor

LI Yong SHAO Cheng

(*Institute of Advanced Control Technology, Dalian University of Technology, Dalian 116024*)
(E-mail: Leeyon76@sohu.com)

Abstract As an effective technique of on-line measurement, soft sensor has been studied by many efforts. Being the premise of soft sensor, it is important to select secondary variables, which can directly influence the efficiency of soft sensor. Grey relation analysis gives a novel method to dig correlation between variables and is suitable to confirm the correlation degree between little-sample data in uncertain systems. Through a detailed analysis of secondary variables' selection for soft sensor, a uniform incidence degree algorithm is presented based on grey relation theory to efficiently select secondary variables. And a practical method to eliminate redundant variables is also proposed. Meanwhile, a practical example is given to verify the feasibility and effectiveness of the method. Simulation results show that the method is efficient and helpful to enhance the accuracy of soft sensor.

Key words Soft sensor, grey relation analysis, uniform incidence degree, radiate basis network

1) 国家科技攻关计划 (2001BA204B01) 和辽宁省科技攻关计划资助
Supported by State Science & Technology Pursuing Project of P.R.China (2001BA204B01) and Science & Technology Pursuing Project of Liaoning Province
收稿日期 2005-4-12 收修改稿日期 2005-10-31
Received April 12, 2005; in revised form October 31, 2005

1 引言

无论是过程控制中先进过程控制算法或策略的具体实施, 还是操作优化、生产协调、故障诊断、状态监测等, 其工程实现的前提都是有效获取过程信息. 然而, 过程检测技术的局限, 一定程度上阻碍了有用信息的获得, 这也正是许多先进的控制算法和策略目前只能停留在理论探讨阶段的重要原因.

一般解决工业过程的测量问题有两条途径: 一是沿袭传统的检测技术发展思路, 通过研制新型的过程测量仪表, 以硬件形式实现过程参数的直接在线测量; 另一种就是采用间接测量的思路, 利用易于获取的其它相关测量信息, 通过计算来实现对被测变量的估计. 由于技术或经济的原因, 工业控制中存在大量目前尚难以或无法通过硬件仪表和传感器进行检测的过程变量和参数. 然而, 这些变量和参数往往与产品质量和生产效率密切相关, 需要加以严格控制或及时了解. 作为解决这一问题的有效方法, 软测量技术 (Soft sensor 或 Soft analyser) 近年来发展迅速, 受到了越来越广泛的重视, 已经成为当前自动控制及检测领域发展的主导方向之一, 在过程工业中已经有了许多成功的应用^[1~4]. 软测量技术的基本原理是依据对可测易测过程变量 (称为辅助变量, 如压力、温度等) 与难以直接测量的待测过程变量 (称为主导变量, 如产品分布、物料成分) 之间的数学关系的认识, 采用各种计算方法, 用软件实现辅助变量对主导变量的测量或估计. 目前, 软测量技术在工业过程中主要应用于实时估计, 故障冗余, 智能校正和多路复用等方面^[5].

软测量技术可分为辅助变量选择、数据处理、软测量模型建立和模型校正等四部分. 其中, 选择辅助变量是其它各步骤的基础和实施前提, 合理高效地选择辅助变量可以大大简化数据处理的复杂性、提高建模和校正速度, 使得软测量过程能更好的满足实际控制的需要.

2 辅助变量选择

软测量过程的辅助变量或称二次变量 (Secondary variable) 的选择包括数量、类型和检测点的选择. 一般来说, 原始可测变量数目, 类型很多, 往往有数十个, 并且相关程度差异较大, 为了实时运行方便, 有必要对原始可测变量进行适当的选择, 以达到降维的目的. 而根据灵敏、准确、特异性以及鲁棒性强的原则, 常用的选择方法有两种: 一种方法是通过机理分析的方法, 找到那些对主导变量影响大的相关变量; 另一种方法是采用主元分析 (PCA)、部分最小二乘法 (PLS) 等统计方法^[6,7] 进行数据相关性分析, 剔除冗余的变量, 降低系统的维数.

然而, 在实际工业生产过程中, 受控对象大都是灰箱系统, 由于很难获得完整全面的机理知识, 因此, 由机理分析得到的对主导变量具有一定影响的变量仍是大量的, 这就导致了基于机理分析的辅助变量选择方法由于考虑因素过多难以在实际中广泛应用. 同时, 由于复杂的生产过程中存在大量可测变量, 单纯的应用 PCA 或 PLS 方法, 会引入高维矩阵运算和繁琐的显著性检验, 大大增加计算量, 降低变量筛选速度, 且经过计算所得的变量集往往仍十分庞大. 因此, 单一的应用以上的两类方法很难满足实际需要.

较为实用的是将二者结合起来, 首先进行粗选, 利用机理分析或经验知识确定一些变量作为预选变量, 在这里, 也可以应用数据挖掘技术对大量的历史数据进行分析^[8], 提取具有希望相关模式的变量作为预选变量. 然后进行细选, 应用适当的算法, 从经过规模缩简的预选变量集中选取需要的辅助变量. 遗憾的是, 细选过程中, 常用的 PCA 和 PLS 算法存在着只适用于大样本过程、算法复杂、计算量大的缺点, 同时, PCA 算法在选取主

元时以最大限度地概括原自变量空间的数据变化信息为准则, 因此保留的主元很可能会夹杂进一些对于回归无益的噪声; 而 PLS 算法在解决非线性问题时, 其内部回归模型的建立方法还有待完善. 此外, 经过 PCA 和 PLS 计算所得的各变量已不具有明确的物理意义, 不易于理解和解释^[5]. 基于以上原因, 本文结合经验知识, 尝试性地将灰关联分析方法应用到了软测量辅助变量的细选过程中.

3 灰关联分析

灰色关联分析是灰色理论^[9]的基本内容, 它与传统的相似性分析方法, 如相关系数法、绝对减数法等^[10]相比, 具有简单、直观和计算量小的优点. 同时, 它对样本量的大小没有过高的要求, 不需要典型的分布规律, 而且其定量分析的结果与定性分析的结果一般能够吻合, 所以, 在对存在不确定性知识的灰色系统进行分析时, 它具有不可比拟的优越性能. 灰关联分析的基本思想是根据数据曲线间相似程度来判断因素间的灰关联度. 灰关联度描述了系统发展过程中, 因素间相对变化的情况, 也就是变化大小、方向与速度的相关性. 如果两者在发展过程中, 相对变化趋势存在相关性, 则可用一定的关联度来表示两者之间的关系. 根据序列之间的关联度的大小, 可以定量判断两序列之间的关联程度, 而且, 应用关联矩阵还可进行多因素情况下的优势分析. 目前, 关于关联度的计算方法很多, 如邓氏关联度^[9]、斜率关联度^[11]、T 型关联度^[12]、C 型关联度^[13]等. 它们已广泛应用于工业控制、农业、管理、地质等领域.

然而, 许多传统的关联度分析算法, 由于存在着一些缺陷^[11~15]而无法在更宽的领域内得到应用. 例如最早应用的邓氏关联度, 它可以对相对平滑的序列间的关联度作出准确的界定, 但由于它在分析过程中只考虑了两序列之间各点距离的大小, 对序列摆动和变化趋势问题没有考虑, 所以在对振荡和分布复杂的数据进行分析时, 它的结果往往与实际不相符.

相似关联度法^[15]是近年来提出的又一种用于因素分析的灰关联算法, 可以一定程度的应用于实际. 但由于它存在零关联度归属不明确的问题, 在一些情况下会导致分析结果失真, 因此, 它的应用范围受到了限制.

此外, 文献 [16] 提出的点关联度算法能对变量的同向相关性作出正确判断, 却无法对工业实际中普遍存在的变量负相关性给出合理的结果.

更值得注意的是, 现有的一些关联度算法存在一个普遍的问题, 即它们只单纯的根据各数据序列中对应的相邻两点间变化量来计算各变量的相关性, 而没有考虑各变量变化率的差异和变化对相关性的影响. 文献 [13,17] 给出的方法虽然考虑了变化率对关联度的影响, 但它们没有给出变化率因子对关联度有效的作用方式, 只是简单的将其与变化趋势关联度进行叠加, 没有充分的利用变量变化率中潜在的信息, 以至在处理单调变化数据与振荡数据的相关性时会出现较大的偏差. 同时, 根据该两种方法的定义, 它们缺少适当的数据无量纲化方法.

为有效解决上述方法存在的问题, 本文在现有的关联度方法的基础上, 提出了一种可根据变量之间的发展态势的相异或相似程度以及变量变化率的差异, 来全面分析序列之间的潜在相关程度的计算方法——一致关联度法. 其算法如下

$$\mu_{ij}(k) = \frac{1}{1 + \left| \frac{|\Delta x_i(k)|}{\Delta_i} - \frac{|\Delta x_j(k)|}{\Delta_j} \right|} \quad (1)$$

$$\Delta x_i(k) = x_i(k+1) - x_i(k) \quad (2)$$

$$\Delta x_j(k) = x_j(k+1) - x_j(k) \quad (3)$$

$$\bar{\Delta}_z = \frac{1}{m-1} \sum_{k=2}^m |\Delta x_z(k)|, \quad z = i, j \quad (4)$$

其中: x_i, x_j 代表经过初始化处理的不可测变量和可测变量, k 为序列编号, m 为样本数, μ_{ij} 为两变量的相关系数, α 为分辨系数, r_{ij} 为关联度.

定义 ξ_k 为符号因子, 则

若 $\Delta x_i(k)\Delta x_j(k) > 0$ 或 $\Delta x_i(k) = \Delta x_j(k) = 0$, 则称 x_i, x_j 在 k 点处趋势相同, $\xi_k = 1$, 关联度为正;

若 $\Delta x_i(k)\Delta x_j(k) = 0$ 且 $\Delta x_i(k) \neq \Delta x_j(k)$, 则称 x_i, x_j 在点 k 处趋势无关, $\xi_k = 0$, 对关联度无贡献;

若 $\Delta x_i(k)\Delta x_j(k) < 0$, 则称 x_i, x_j 在点 k 处趋势相反, $\xi_k = -1$, 关联度为负.

$$v_{ij}(k) = \frac{1}{1 + \left| \frac{|\Delta x_i(k) - \Delta x_i(k-1)|}{\bar{\Delta}_i} - \frac{|\Delta x_j(k) - \Delta x_j(k-1)|}{\bar{\Delta}_j} \right|} \quad (5)$$

$$r_{ij} = \left| \frac{1}{m-1} \sum_{k=1}^{m-1} \xi_k \mu_{ij}(k) \right|^\beta, \quad \beta = \left| \frac{1}{m-2} \sum_{k=2}^{m-1} \xi_k v_{ij}(k) \right| \quad (6)$$

其中, 等式 (6) 中指数 β 代表数据的变化率对关联度的影响因子. 当变量间变化率的差异单向增加时, 其作用尤其明显; $v_{ij}(k)$ 为变化率关联系数.

设两变量序列 x_i, x_j 有 m_1 个趋势相同点, 编号为 $\{t_1 \cdots t_{m_1}\}$; m_2 个趋势无关点, 编号为 $\{s_1 \cdots s_{m_2}\}$; m_3 个趋势相反点, 编号为 $\{d_1 \cdots d_{m_3}\}$; 且

$$P_{ij} = \frac{1}{m-1} \sum_{k_1=t_1}^{l_{m_1}} \xi_{k_1} \mu_{ij}(k_1) \quad (7)$$

$$Z_{ij} = \frac{1}{m-1} \sum_{k_2=s_1}^{s_{m_2}} \xi_{k_2} \mu_{ij}(k_2) \quad (8)$$

$$N_{ij} = \frac{1}{m-1} \sum_{k_3=d_1}^{d_{m_3}} \xi_{k_3} \mu_{ij}(k_3) \quad (9)$$

$$m_1 + m_2 + m_3 = m - 1 \quad (10)$$

$$r_{ij} \left| \frac{1}{m-1} \sum_{k=1}^{m-1} \xi_k \mu_{ij}(k) \right|^\beta = |P_{ij} + N_{ij}|^\beta \quad (11)$$

P_{ij}, Z_{ij}, N_{ij} 分别代表正关联度, 零关联度, 负关联度.

当 $P_{ij} > |N_{ij}|$ 时, 两变量以正相关为主, 即它们的变化趋势相似, 相关程度由 P_{ij} 及 r_{ij} 两因素的大小来衡量; 当 $r_{ij} = Z_{ij} = 0$ 时, 两变量无关; 当 $P_{ij} < |N_{ij}|$ 时, 两变量以负相关为主, 即它们的变化趋势相反, 相关程度由 $|N_{ij}|$ 及 r_{ij} 两因素的大小来衡量. 应注意的, 在估计与预测中, 负关联也是变量之间相关的表现, 通过建立适当的模型, 负关联的变量可以相互表征.

一致关联度不仅满足灰关联分析的基本性质:

- 1) 规范性, 即 $r_{ij} \in [0, 1]$, 且当两变量全相关时 $r_{ij} = 1$. 证明略.
- 2) 偶对称性, 即 $r_{ij} = r_{ji} \Leftrightarrow x = \{x_i, x_j\}$, 显然.

3. 保序性, 即在无量纲处理后, 序列数据的大小顺序不变. 即当 $x_z(k) < x_z(k+1)$ 时, 有 $\frac{x_z(k)}{\Delta_z} < \frac{x_z(k+1)}{\Delta_z}$. 而且, 它具有一些新的优点

1) 从实际出发, 明确了零关联度的物理意义.

2) 以数据变化量均值化代替了传统的初值化、均值化和区间化^[17] 等数据无量纲化处理. 初值化削弱了线性相关变量的相关性. 同时, 原来两个序列发展态势相同, 经初值化后, 初值大的发展态势变慢了, 初值小的发展态势相对增大了, 导致了数据的特征失真; 均值化可以一定程度上缓和初值化的缺点, 但不能从根本上保证数据的特征的完整, 同时, 在数据均值为零时, 它会出现被除数为零的情况; 区间化可以保证数据变化趋势不变, 但经区间化后, 很明显变量间关联系数的值域变为 $[0.5, 1]$, 描述宽度只有其它方法的一半, 分辨率较低. 而由定义 (1)~(5) 可知, 一致关联度算法可以较好的解决上述问题, 在准确反映变量关系的同时, 还具有较好的分辨率.

3) 更符合实际, 同时考虑了数据的变化趋势和变化率对关联度的影响.

4 仿真实验

在选矿行业的磨矿过程中, 磨矿浓度是球磨机运行中的一项重要内部指标, 它对磨矿过程的产量、出口粒度有着重要的影响. 对磨矿浓度进行优化控制, 将大大改善球磨机的工作效率, 提高磨矿过程达标产品的产量, 降低损耗. 然而, 由于磨矿作业影响因素多, 机理复杂, 属于多变量强耦合系统, 目前尚无有效的手段对球磨机磨矿浓度进行直接在线检测. 针对这一问题, 本文作者提出了将采用一致关联度的软测量技术应用于球磨机磨矿浓度的在线检测, 并以此检验一致关联度方法的有效性.

由于球磨机属于部分信息可知的灰色系统^[18], 由先验知识可知, 球磨机磨矿浓度与一些外部变量如给矿量、排矿水量、声响、给矿粒度等, 存在一定相关性. 根据软测量思想, 由于大部分磨机外部变量是在线可测的, 可以通过分析它们与磨矿浓度的潜在相关性的强弱, 选择出关联性强外部变量对磨矿浓度进行实时估计. 因此, 结合先验信息, 采用一致关联度对磨矿浓度及与其存在一定关联性的九项可测变量进行分析, 结果如表 1.

表 1 外部参数与磨矿浓度的关联度

Table 1 Incidence degree between external parameters and grinding concentration

与磨矿浓度的关联度		与磨矿浓度的关联度	
给矿量	0.312	输入功率	0.761
返砂水量	0.133	排矿水量	0.029
声响	0.681	溢流浓度	0.426
给矿粒度	0.212	磨机转速	0.707
轴承压力	0.832		

通过相关性分析可知, 在指定满意阈值为 0.7 的前提下, 在由先验信息确定的预选变量中, 球磨机外部参数: 压力、声响、输入功率、球磨机转速率与待估磨矿浓度的关联度满足要求. 进一步, 根据辅助变量选择原则, 在保证与主导变量具有较强相关性的同时, 还要考虑到各辅助变量的独立性, 剔除冗余变量, 简化建模的复杂程度. 然而, 对于强耦合非线性系统, 变量的冗余是相对的, 很难找到完全线性相关的变量, 因此, 冗余变量剔除的问题就变成了协调估计精度和预测模型复杂程度的综合优化问题, 即在保证满意的估计精度的前提下最大程度地剔除冗余变量. 基于这一思想, 作者给出一种实用的冗余变量剔除方法, 步骤如下

1) 计算各辅助变量间的相关性.

2) 设定一阈值, 将相关性大于该值的变量定义为相似变量, 它们组成的集合定义为 C , 剩余变量的集合定义为 F , 设 C 中共有 m 对 n 个不同的相似变量 ($2m \geq n$).

3) 将 C 中每对相似变量中与主导变量关联性强的一个放入 F (如某变量被重复选中, 则只算一次), 设共有 t 个变量被选中, 它们与 F 中原有的变量组成 F^+ , 则 F^+ 中的变量为固定参与建模的变量.

4) 按组合规则从 C 中剩余的变量中分别选取 0 到 $n-t$ 个变量与 F^+ 中的变量共同构成不同的辅助变量组, 分别建立软测量模型.

5) 比较各模型的综合性能, 选择估计精度满足要求, 变量个数最少的一个.

设定阈值目的是预先保留一部分相关性小的变量, 可以起到减少上述过程的计算量的作用. 当先验知识不足时, 选取的阈值可能偏大, 达不到剔除冗余变量的目的, 此时可以对阈值作适当调整.

本文中, 由于原始辅助变量个数较少, 作者对所有的四种分类情况都进行了计算, 即阈值设为零. 采用 RBF 神经网络进行软测量建模, 采用 OLS 算法^[19] 进行网络训练.

经上述变量集优化得, 当四变量都参与建模时, 所建模型的估计精度可以满足实际要求, 其所建立的模型性能如表 2 所示.

表 2 实验性能指标
Table 2 Experimental performance index

训练数据	测试数据	训练相对误差平均值	训练相对误差最大值	测试相对误差平均值	测试相对误差最大值
50 组	14 组	0.21%	0.3%	0.64%	0.91%

其中由测试数据所得的磨矿浓度估计曲线如图 1.

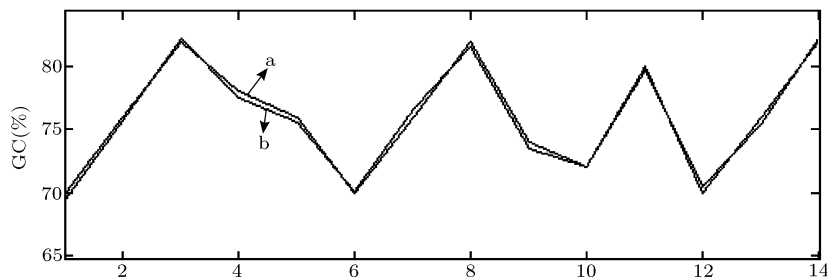


图 1 磨矿浓度软测量结果 (a: 经数据处理后的实际值; b: 估计值)

Fig. 1 Soft sensor result of grinding concentration (a: Processed real value; b: Estimation)

5 结论

通过理论分析, 一致关联度克服了以往关联度算法所存在的问题, 能得出更符合实际的分析结果. 同时, 仿真实验验证了: 应用一致关联度法对软测量辅助变量进行选择, 可以快速准确的提取变量间的相关性, 大大降低运算量, 提高建模速度. 建立的软测量模型具有较高的精度, 预测平均相对误差保持在 0.7% 以下, 可以满足实际控制应用的要求. 此外, 本文将灰关联分析方法应用于软测量, 在一定程度上解决了以往机理分析等方法存在的问题, 为拓宽软测量技术的应用范围, 增强其有效性提供了一种有效手段.

References

- 1 Mejdell T, Skogestad S. Output estimation using multiple secondary measurements: High-purity distillation. *American Institute of Chemical Engineers Journal*, 1993, **39**(10): 1641~1653
- 2 Park S, Han C. A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns. *Computers and Chemical Engineering*, 2000, **24**: 871~877
- 3 Albertos P, Goodwin G C. Virtual sensors for control applications. *Annual Reviews in Control*, 2002, **26**: 101~112
- 4 Wang X H, Li S Y, Xi Y G. Research on soft sensor model for slab temperature in reheating furnace. *Acta Automatic Sinica*, 2004, **30**(6): 928~932
- 5 Yu J S, Liu A L, Zhang K J. Soft Sensor Technology and Its Application in Petrochemical Industry. Beijing: Chemical Industry Press, 2000. 56~102
- 6 Geladi P, Kowalski B. Partial least squares regression: A tutorial. *Analytical Chemica Acta*, 1986, **185**: 1~17
- 7 Liu R L, Su H Y, Zhu J. Application of soft sensing method based on PLS regression for estimating the stabilizing oil components of catalytic reform reactor. *Control and Instruments in Chemical Industry*, 2002, **29**(5): 44~47
- 8 Chen L D, Toru S. Data mining methods, applications, and tools. *Information System Management*, 2000, **17**(1): 65~70
- 9 Deng J L. Grey Theory Basis. the 2nd edition. Wuhan: Huazhong University of Science and Technology Press, 2002. 122~154
- 10 Yang G B, Gao Y Y. Fuzzy Mathematics Theory and Application. Guangzhou: South China University of Technology Press, 2001. 115~125
- 11 Li X Q. Research on the computation model of grey interconnect degree. *Systems Engineering*, 1995, **13**(6): 58~61
- 12 Tang W X. The concept and the computation method of T's correlation degree. *Mathematics Statistics and Management*, 1995, **14**(1): 34~38
- 13 Wang Q Y, Zhao X H. The relational analysis of C-mode. *Journal of Huazhong University of Science and Technology*, 1999, **27**(3): 75~77
- 14 Li H Y. Study on the calculation method of grey relationship degree. *System Engineering and Electronics*, 2004, **26**(9): 1231~1234
- 15 Ma B G, Cheng G Q. A formula of similarity correlation degree. *Systems Engineering-Theory & Practice*, 2000, **7**: 69~71
- 16 Xu X L. Improvement and Application of Grey-Relation, -Clustering, and -Prediction. [Master Degree Dissertation], Xiamen: Xiamen University, 2001. **6**: 17~22
- 17 Wang Q Y. The grey relational analysis of B-mode. *Journal of Huazhong University of Science and Technology*, 1989, **17**(6): 77~82
- 18 Wang Z H. Study on Measurement Optimization of the Internal Parameters in Ball Mill. [Ph. D. Dissertation], Shenyang: Northeastern University, 2001. **7**: 19~25
- 19 Chen S, Cowan C F N, Grant P M. Orthogonal least squares learning algorithm for radiate basis function network. *IEEE Transactions on Neural Network*, 1991, **2**(2): 302~309

李 勇 现就读于大连理工大学控制理论与控制工程专业, 博士研究生, 从事工业过程先进控制技术及其应用的研究。

(**LI Yong** Ph.D. candidate in the Department of Control Theory and Control Engineering at Dalian University of Technology. His research interests include advanced control technology of industry process and its application.)

邵 诚 1992 年毕业于东北大学工业自动化专业, 现为大连理工大学先进控制技术研究所所长, 教授, 博士生导师, 研究领域为复杂系统的建模与控制, 自适应控制, 鲁棒控制等。

(**SHAO Cheng** Graduated from Industry Automation Department of Northeastern University in 1992. He is a professor in Institute of Advanced Control Technology at Dalian University of Technology. His research interests include modeling and control of complex system, adaptive control, and robust control.)