

一个测量装置在大规模制造中的标定问题

基本条件和实验要求 (V2.2 2021-2-17)

| | |
|-------------------------|---|
| 1. 概念词汇说明 | 1 |
| 2. 符号 | 2 |
| 3. 工程问题背景 | 2 |
| 4. 建立数学模型 | 3 |
| 4.1 装置的原理框图 | 3 |
| 4.2 传感器部件的输入-输出特性 | 3 |
| 4.3 标定工序 | 3 |
| 4.4 标定方案的评估比较 | 4 |
| 5. 样本数据集电子文件和格式 | 5 |
| 6. 课题研究要求 | 5 |
| 7. 报告写作 | 5 |
| 7.1 写作的基本要求 | 5 |
| 7.2 格式参考 | 6 |
| 8. 课题研究报告评判要领 | 6 |
| 9. 避免不正当使用他人工作成果 | 6 |

1. 概念词汇说明

拟合 Curve Fitting

科学或工程上可以通过实验等方法获得关于某个问题的若干离散数据。依靠数学方法，使用更为密集离散方程或者连续函数（曲线方程）尽量逼近这些已知离散数据点集，此过程称为拟合。一般使用最小二乘准则作为逼近误差程度的判则。

插值（内插） Interpolation

一般是指通过函数在有限个离散点处的取值状况，估算出函数在其他点或区域处的近似值。通常，这种估算会借助曲线拟合得出的表达式来进行。

标定 Calibration

确定仪器或测量装置的输入-输出关系，赋予其分度值。

搜索算法 Algorithm of Search

利用计算机的高性能运算能力，有目的地检验某个问题的解空间的局部或全部，从而求得问题的优化解的做法。

启发式搜索 Heuristic Search

在搜索问题解空间时，对当前已搜索的区域进行评估和线索分析，寻求下一步最优搜索方向，以便减小总体搜索范围，降低求解问题的复杂度。在某些情况下，简单穷举搜索会遭遇“组合爆炸”窘境，启发式搜索有助于应对此类情况。

课题自定义词汇：单点测定、单点测定成本

在本课题中，当被测温度 T 为任一定值时，对传感器输出电压 V 进行观测和记录的过程，称为一次单点测定。完成单点测定要付出一定成本，称为单点测定成本。

课题自定义词汇：标定误差成本

在本课题中，测量装置完成标定后，按给定规则，被测温度 T 取一系列指定值时，将装置的测量误差折算为一种成本，称为标定误差成本。

课题自定义词汇：样本数据集、训练用样本数据集、测试用样本数据集

课题先期已对随机选取的传感器个体样品做了较细致精确的测试，所得实验数据汇总成表，供课题研究之用，称为样本数据集。按使用场合分成训练用样本数据集（简称训练集）、测试用样本数据集（简称测试集）。

训练集供学习者作为课题研究线索，用于试验构建数学模型，尝试寻找或训练求解算法。

测试集用于检验最终选择的模型的性能。测试集又分为测试集 A 和测试集 B，前者供学习者使用，后者对学习者可不可见，供教师做评价验证和打分时使用。

上述各样本数据集具有同源性。

2. 符号

| | |
|-----------------|--------------------------------------|
| V | 传感器的输出电压 |
| T | 传感器的输入，即被测温度 |
| \hat{T} | 测量装置的输出（读数），是测量装置对输入信号 T 的估计值 |
| $T_{i,j}$ | 第 i 个样本之第 j 点对应的实际温度值（来自标准样本数据库） |
| $\hat{T}_{i,j}$ | 第 i 个样本之第 j 点对应的温度估测值 |
| $s_{i,j}$ | 第 i 个样本之第 j 点对应的单点误差对应的成本数值 |
| Q | 单点测定成本 |
| n_i | 对样本 i 标定过程中的测定点数目 |
| S_i | 对样本 i 的标定成本 |
| M | 标准样本数据库中的样本总数 |
| C | 基于标准样本数据库评价一个标定方案，算得的该方案总体成本 |

3. 工程问题背景

很多大规模制造的电子产品中包含有带测量功能的模块（测量装置），用于监测某种物理量，比如环境温度、压力或光照强度等。任何测量装置在制造时一般都要经过标定。

本课题假定一种测量装置。为便于理解，设定被测物理量为环境温度，可测范围是 -20°C 至 70°C 。所用核心传感器的输入-输出特性有明显非线性，且个体差异性比较明显。课题要求为该装置设计一种标定工序的方案，适合大规模高效率批量制造。

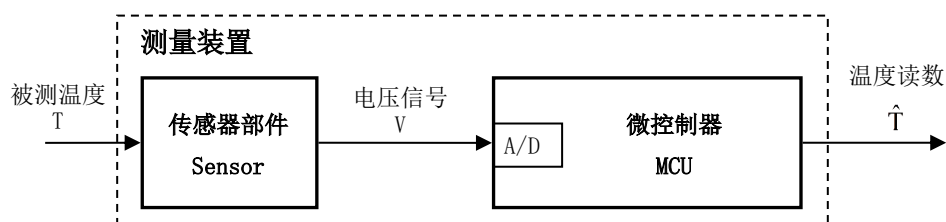


图 1 装置原理框图

4. 建立数字化描述

为适合开展研讨，对问题的一些方面先进行必要的数字化描述，以方便进一步把工程问题转化为数学问题并求解。

4.1 装置的原理框图

测量装置的原理框图见图 1。其中，传感器部件（包含传感器元件及必要的放大/调理电路）的输入信号，即被测温度，以符号 T 表示；它的输出信号是电压形式，用符号 V 表示。该电压信号经模数转换器（ADC）转为数字编码，被微处理器读取和程序处理，获得温度读数 \hat{T} 。所以，微处理器通过检测传感器电压信号，间接计算被测温度。这一操作须基于 $V-\hat{T}$ 函数关系。

对该装置而言，标定就是指确定函数关系 $\hat{T}=f(V)$ 的过程。该函数关系与传感器部件输入-输出关系特性 $T-V$ 密切相关。

4.2 传感器部件的输入-输出特性

根据课题设定，测量装置内部温度传感器的输入-输出关系特性不是理想线性的，具体表现在较明显非线性和个体差异性。

4.2.1 非线性

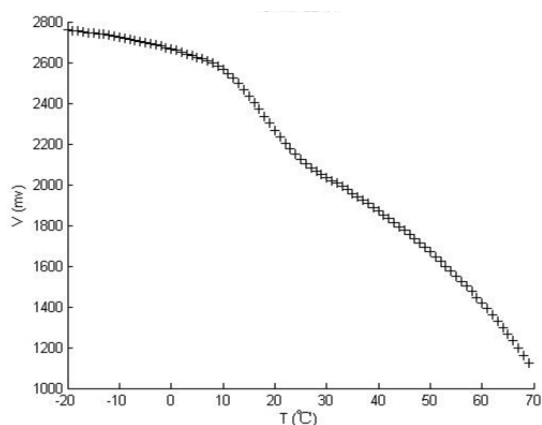


图 2 传感器特性的非线性

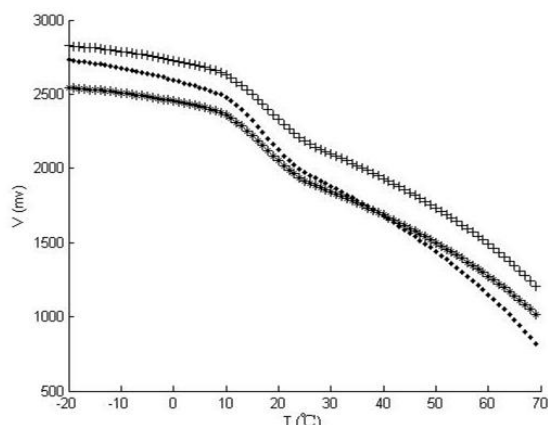


图 3 传感器特性的个体差异

图 2 是对某一传感器个体的测试结果。在可测温度范围内选择测试了 90 个温度点，被测温度分别为 -20 、 -19 、 -18 、...、 67 、 68 、 $69(^{\circ}\text{C})$ ，以符号“+”标示于直角坐标图中。

可见，随着被测温度值上升，传感器输出电压单调地以明显非线性的形态下降。

4.2.2 个体差异性

对不同传感器个体进行测试，发现个体间差异比较明显。图 3 在同一坐标图中绘出三件传感器个体特性测试结果，分别以“+”、“*”、“.”表示。

更多的测量试验结果表明，传感器特性数据在统计上呈现一定随机性。

4.3 标定工序

图 2 示意了对某一传感器个体特性精细测试，获得了关于该个体比较全面的输入-输出情况，共检测了 90 对 (T,V) 数值。理论上，通过拟合等方法，可确定函数关系 $V=g(T)$ ，利用其反函数，令

$\hat{T} = g^{-1}(V)$ ，就完成了对测量装置的标定。但是，在大规模制造中，若每一件产品个体都这样做 90 个温度点测试，则需要大量设备投资，且会降低生产速度，造成生产成本显著过高。

图 4 示意了在自动生产流水线上实施 6 个温度测试点的标定工序运转形态。

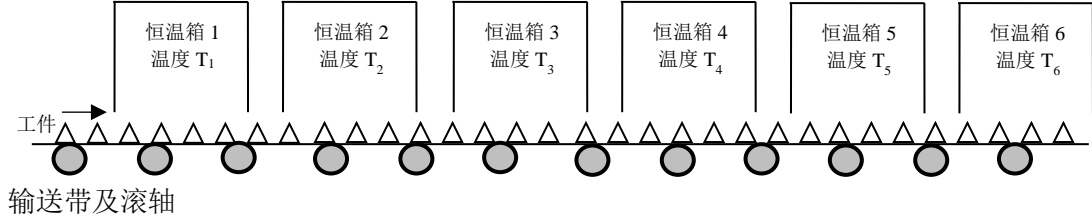


图 4 流水线上的标定工序示意

本课题聚焦于研讨能否减少测试点并获得优化的标定方案。当然，直观上不难理解，作为代价，较少的测试点会降低标定准确度，扩大测量装置的系统误差。优化方案的关键在于做好这两个对立方面的平衡。

4.4 标定方案的评估比较

为评估和比较不同的标定方案，特制定以下成本计算规则。总体成本较低的标定方案，可认定为较优方案。

- 单点测定成本

实施一次单点测定的成本记为符号 Q 。本课题指定 $Q=50$ 。

- 标定误差成本

$$s_{i,j} = \begin{cases} 0 & \text{if } |\hat{T}_{i,j} - T_{i,j}| \leq 0.5 \\ 1 & \text{if } 0.5 < |\hat{T}_{i,j} - T_{i,j}| \leq 1.0 \\ 6 & \text{if } 1.0 < |\hat{T}_{i,j} - T_{i,j}| \leq 1.5 \\ 20 & \text{if } 1.5 < |\hat{T}_{i,j} - T_{i,j}| \leq 2.0 \\ 10000 & \text{if } |\hat{T}_{i,j} - T_{i,j}| > 2.0 \end{cases} \quad (1)$$

单点误差对应的成本值按式（1）计算，以符号 $s_{i,j}$ 记。其中 $T_{i,j}$ 表示第 i 个样本之第 j 点的实际温度值， $\hat{T}_{i,j}$ 表示标定后得到的估计值（读数）。

单个样本个体的标定误差成本用和式（2）计算。

$$S_i = \sum_{j=1}^{90} s_{i,j} \quad (2)$$

- 样本个体标定成本

$$c_i = S_i + Q \cdot n_i \quad (3)$$

样本个体的标定成本是测定成本与误差成本之和，见式（3）。式中 n_i 表示对该样本个体标定过程中的测定点数目。

- 方案总体成本

标定方案的总体成本值按式（4）计算。对选定的样本数据集里的每个样本个体逐一开展标定，取所有样本个体标定成本的统计平均。其中 M 是该数据集样本总数。

$$C = \frac{1}{M} \sum_{i=1}^M c_i \quad (4)$$

5. 样本数据集电子文件和格式

为开展课题研究，先期已对一定数量且足够多的传感器个体样品（样本）做了较细致精确的测试，所得实验数据汇总成表，制作成 csv 格式电子表格文件。该电子表格中，第 $2i - 1$ 行（奇数行）存放第 i 件样本实验被测温度 T 的数值，默认物理单位是摄氏度；第 $2i$ 行（偶数行）存放对应的电压信号 V 的数值，默认单位毫伏。

dataform_train.csv 为训练集的数据文件，dataform_testA.csv 为测试集 A。

6. 课题研究要求

对以上课题开展研究，必要时使用计算机编程辅助分析求解，指定使用 MATLAB 作为工具程序语言。针对以下要求，设计行之有效的算法和工作步骤：

(1) 基本问题研讨

先完成基本问题研讨。将工程问题提炼转化为一个组合优化数学问题。在构建数学模型和求解算法时，指定要使用三次样条插值法和遗传算法；遗传算法编程实现时，不可使用 MATLAB 遗传算法工具箱，程序代码需自编。

(2) 拓展问题研讨

若条件允许，学习者可围绕课题开展自主拓展研讨，研讨范围不受基本要求的约束，比如可以讨论不同的拟合插值方法的效果、不同搜索算法的效果，等等。

7. 报告写作

将课题研究的结果编写成报告，又称小论文，既是学生课程学习成果的总结，也是本课程评分的主要项目。本课题的报告有两次提交。先提交初稿，经过老师评判，给出反馈意见，配合师生交流指导，学习者可学习如何改进和完善，再提交最终正式稿。

7.1 写作的基本要求

- 真实反映试验结果和工作成果；
- 参照专业学术类期刊的文稿样式和学术规范，尽量做好“言之成理，言之有据”；
- 小论文主题目（注：不是指电子文件名称）指定为“一个测量装置在大规模制造中的标定问题”；
- 无须详细介绍课题背景，可假定读者已阅读过《基本条件和实验要求》（即本文）；
- 正文 5 号字体，A4 纸张，允许双列排版，内部标题自拟；
- 核心内容应包括数学模型和求解算法的构建，简洁且全面地展示必要的数学表达、算法结构和步骤、实验结果，说明方案和结果的有效性；
- 如果有自主选做的拓展研讨话题，请在正文中设立醒目的“一级标题”，并将相应内容置于该标题之下；
- 正文篇幅（不计正文前的“摘要”、“名词定义”和文后附录）有限制；不包含自主拓展问题讨

论的报告，正文篇幅以 8 页为上限；包含自主拓展问题的，篇幅可放宽到最多不超过 11 页；

- 需将程序代码完整清单列于文后附录中，不要列在正文中；
- 初稿和正式稿的写作基本要求一致。

7.2 格式参考

提供以下材料作为小论文格式风格的参考。

(1) 发表在某核心科技期刊的专业学术论文《DS 证据理论在雷达体制识别中的应用》

其行文风格、内容编排组织等方面可供借鉴。该文是关于 DS 证据理论的特定应用，而本课题把统计推断、GA、SA 等算法应用于特定问题，有类似之处。

(2) 本课程小论文模板

改编自上海交通大学本科毕业设计论文模板。论文题目请自行改为前文规定，内部各级标题可遵照规则自改，英文摘要为选做内容。

8. 课题研究报告评判要领

表 1 小论文评判要领

| 评判项目 | 要领 | 备注 |
|---------|--|-----------------|
| 数学模型合理性 | 数学逻辑自洽，前提假设合理 | —— |
| 算法说明合理性 | 算法流程清晰 | —— |
| 实验结果合理性 | 与物理实际相一致 | —— |
| 结果分析合理性 | 从实验数据和现象中得到合理和有实用意义的分析结果 | —— |
| 报告整体完整度 | 标题、摘要、引言、正文、附录等 | —— |
| 格式规范程度 | 排版、图表、公式等 | —— |
| 学术规范程度 | 引文注解，禁止抄袭 | 实质性抄袭可能直接导致不及格！ |
| 独立性 | 观点、方法的独特性，对研究内容的自主拓展等 | —— |
| 总评 | 设等级 A+, A, A-, B++, B+, B, B-, B, C+, C, C-, D, F F 对应不及格 | |

9. 避免不正当使用他人工作成果

科学研究中，经常有必要借鉴和学习一点他人的工作和成果，可以获得更高的工作起点，可以少走弯路。但是，我们在发表自己的工作成果时，必须遵循学术规范和惯例做法，把自己的工作中哪些地方是引用或参考他人的工作，哪些地方是本人的独立原创工作，加以明确表述，尽量避免引起读者误解，防止构成事实上的不正当使用他人成果。

引用和借鉴他人的工作成果时，只要严格按照学术规范做，就是完全正当的。

在本课程中，如果你借鉴了同届或往届同学的工作，包括但不限于文字段落、图、表、计算机程序代码等，都必须将相关报告列入小论文参考文献列表，并按引文规范行事。

同样，报告中引用本课程的讲义或教学材料的内容时，最好也要添加引文注解。

你的小论文中任何未加引用注释或表述的非原创内容，都可能使评判老师误解为作者原创，导致评分时出现误判。此类不遵守规范的情况一经发现，将给予重罚。