

# Multi-Sensor Information Fusion Technology Based on Projection Pursuit Simulated Annealing

Youhao Zhang, Yanfang Zheng\*

School of Electrical and Information Engineering, Zhangjiagang Campus, Jiangsu University of Science and Technology, Suzhou Jiangsu

Email: 18724145220@163.com, \*zyf062856@163.com

Received: Mar. 9<sup>th</sup>, 2020; accepted: Mar. 24<sup>th</sup>, 2020; published: Mar. 31<sup>st</sup>, 2020

---

## Abstract

Based on the advantages of weighted average method and probability statistics method, a multi-sensor information fusion algorithm based on projection pursuit simulated annealing method is designed in this paper. The innovation of the design lies in the addition of probability window constraint and clustering constraint in the process of projection weight adjustment to reduce the computational complexity of the system by 41.6%, and the results obtained are in line with the peak regression judgment. This design has a new improvement to solve the multi-sensor information fusion of the system, and has a certain improvement to the real-time and applicable range.

## Keywords

Multi-Sensor Information Fusion, Projection Pursuit, Simulated Annealing, K-MEANS

---

# 基于投影寻踪模拟退火的多传感信息融合技术

张友好, 郑艳芳\*

江苏科技大学张家港校区电气与信息工程学院, 江苏 苏州

Email: 18724145220@163.com, \*zyf062856@163.com

收稿日期: 2020年3月9日; 录用日期: 2020年3月24日; 发布日期: 2020年3月31日

---

## 摘要

本文吸取加权平均法、概率统计法等算法的优点, 设计了基于投影寻踪模拟退火法的多传感器信息融合\*通讯作者。

算法。设计创新点在于对投影权值调整过程中加入概率窗口约束及聚类约束, 使系统的计算复杂度减少 41.6%, 得出的结果符合峰值回归判断。本设计对于解决系统的多传感器信息融合有了新的改进, 对于实时性和适用范围均有了一定的提升。

### 关键词

多传感器信息融合, 投影寻踪, 模拟退火法, K-MEANS

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 绪论

5G 时代, 万物互联不再是空话。高达 TB 级的传输速率和海量节点数, 为智能生产生活赋予了强大的助推力。未来, 空间里将覆盖着密集传感器网络, 对信息的把控将更为彻底。许多情况下, 并不需要了解每一个传感器的数据, 更想获取的是区域内的总体指标参数, 简化决策, 实时找到故障区域。为解决这一问题, 多传感器的信息融合技术正在被众多学者研究[1]。但其相关技术、算法, 都有一定的局限性和缺陷。所以, 还需要进一步的探索, 为 5G 时代传感器的改革创新铺设通畅的道路。

多传感器信息融合(Multi-sensor Information Fusion, MSIF), 是利用计算机技术将来自多个传感源的信息和数据, 在一定准则下加以自动分析和综合的信息处理过程。具体流程如图 1 所示。

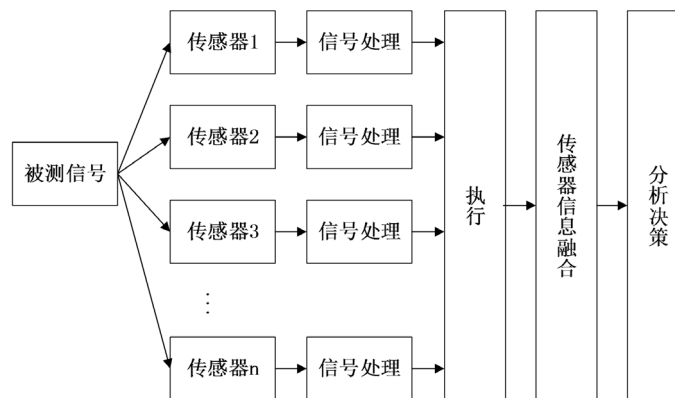


Figure 1. Multi-sensor information fusion process  
图1. 多传感器信息融合流程

目前, 主流的多传感器信息融合技术主要分为三类[3]。第一类为先验知识型, 以加权平均法为代表; 第二类为概率分配型, 包括卡尔曼滤波法、D-S 证据理论; 第三类为训练学习型, 例如模糊逻辑控制、专家系统、人工神经网络法等[2]。各类算法的介绍如下: (1) 加权平均法[4]: 对多个传感器的数据进行加权、平均、累加等算术操作, 最终得到一个统一的指标。该方法操作简单、计算迅速、复杂程度低, 但需预先获取各个传感器的加权值。合理精确的加权值难以获取且随时间、数据变化的关系不明确, 固定的加权值可能会造成较大的误差。(2) 概率论方法[5]: 对多个传感器的数据进行统计分析、去量纲化、剔除奇点等操作, 利用概率论方法, 结合先验知识等。最终得到统一的指标。这种方法理论依据充足、

响应快速, 但需先验知识, 对于黑盒类型系统不适用。(3) D-S 证据理论[6]: 无需先验知识, 而是使用基本的概论分配函数代入合成公式中, 进行信息融合, 是处理不确定问题的通用框架。但是随着信息规模、类型的不断扩大, 计算时间成几何倍数的增长, 实时性差, 不利于实时性的决策。

实际在进行多传感器信息融合的过程中, 出现了许多的问题和难点。具体如下:

- (1) 数据固有缺陷: 不和谐数据奇点、偶然误码。
- (2) 量纲和维度: 量纲、维数不统一。
- (3) 数据关联: 正负性指标需分别处理、数据类型多。

针对现有主流多传感器信息融合技术在适用范围、系统适应度、先验知识依赖强等缺陷[1], 本文设计了一种基于投影寻踪和模拟退火法的多传感器信息融合算法来改善解决这些问题。第 2 部分对该算法进行详细的描述, 并对算法进行逐块的介绍和分析讨论。第 3 部分给出了该算法测试结果, 并将该结果与实际数据对比, 与在未融合信息的传感器数据进行对比。第 4 部分对本研究工作进行了总结与讨论。

## 2. 方法

### 2.1. 前人的方法及改进流程

前人运用了多种类型的算法来解决多传感器信息融合的问题。主要有(1): D-S 证据理论, 传统的 D-S 证据理论框架为利用有限个基本概率分配函数  $m_1, m_2, \dots, m_n$ , 来构造一个辨识框架, 从而进行系统辨识, 给出一个综合指标。D-S 证据的理论合成公式:

$$(m_1 \oplus m_2 \oplus \dots \oplus m_n)(A) = \frac{1}{1-k} \sum_{A_1 \cap A_2 \cap \dots \cap A_n = A} m_1(A_1) m_2(A_2) \dots m_n(A_n)$$

(2) 加权平均法, 加权平均法基本原理是观察大量样本, 取平均权值作为经验权值, 将各个样本数值与权值进行累积求和。基本公式为  $s_j = \sum_{i=1}^n c_{ij} w_{ij}$ , 其中  $w_{ij}$  为样本  $j$  的第  $i$  个指标的经验权值,  $c_{ij}$  为第  $j$  个样本值。(3) 概率论方法, 概率论法是基于频率和权数的转化方法, 在加权平均法基础上进一步作出的改进。研究的思想是某一类信息数据对系统出现较大影响频率作为该类型信息的权重划分依据。基本公式如下:

$$\bar{\mu}_i = \frac{1}{m} \sum_{j=1}^n x_{ij}$$

$$\sigma_i^2 = \frac{1}{m-1} \sum_{j=1}^n (x_{ij} - \bar{\mu}_i)^2$$

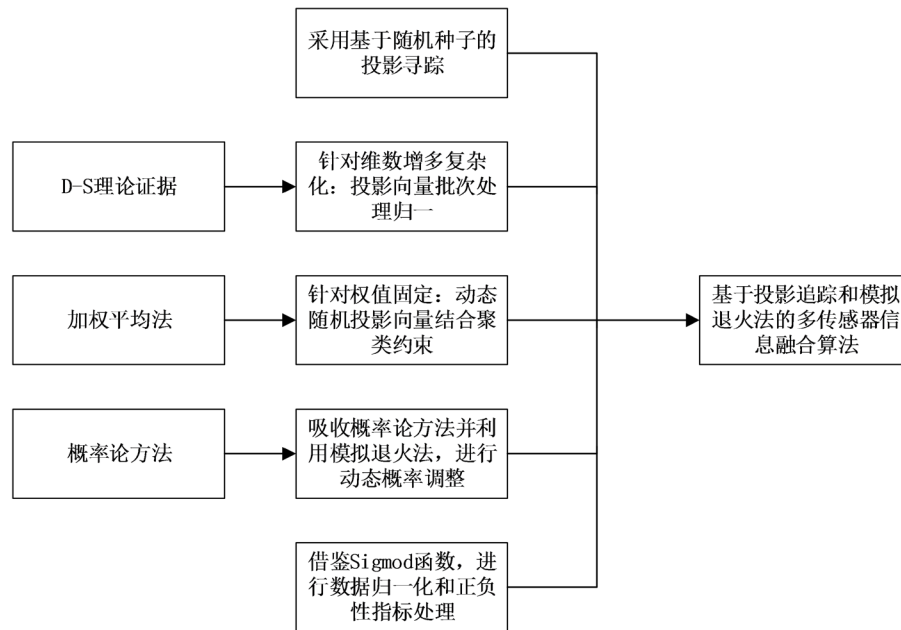
本算法设计基于前人的基础进行了改进, 改进的流程和具体内容如图 2 所示。

改进后的算法具体流程如图 3 所示。

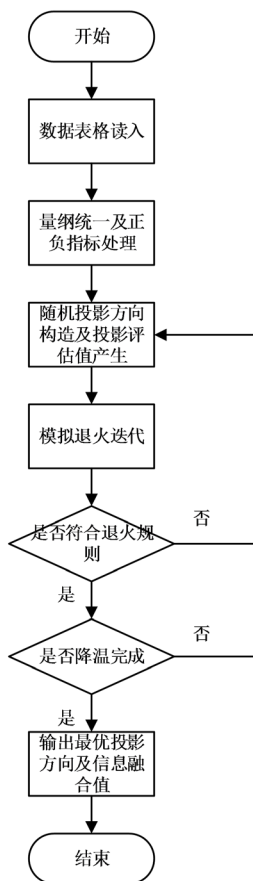
### 2.2. 统一量纲化及正负性指标处理

#### 统一量纲化及正负性指标处理的原理

实际的生产生活问题中, 传感器数据类型丰富。如温度、湿度、压力、pH 值、光照等等。针对某一项指标, 方便统计, 但如果要求根据多个指标给环境或状态一个综合的评估值, 可能比较困难。人们借自身知觉或经验进行直接的感受评估, 但机器没有这种机能。将各种类型丰富的数据去量纲化, 只保留本质的特征因子, 会便于机器的准确计算和决策。对于有些类型的数据值, 希望越大越好, 而有些数据值, 却希望越小越好。这即是正负指标的问题。因此要解决量纲去除的同时还要考虑到正负指标的处理。本文采用的数据处理方法如下:



**Figure 2.** Multi-sensor information fusion algorithm to improve the process  
**图2.** 多传感器信息融合算法改进流程



**Figure 3.** Overall flow chart of fusion algorithm  
**图3.** 融合算法总体流程图

定义某类型数据为  $x_j$ ,  $\max x_j$  为第  $j$  列指标的最大值,  $\min x_j$  为第  $j$  列的最小值。对于正向性指标, 做如下处理:

$$x_j = \frac{x_j - \min x_j}{\max x_j - \min x_j}$$

对于负向性指标, 做如下处理:

$$x_j = \frac{\max x_j - x_j}{\max x_j - \min x_j}$$

## 2.3. 投影寻踪法原理及应用

### 2.3.1. 投影方向的设计

对于未知系统, 没有先验经验, 或者系统为非线性、时变的。投影方向的选择往往也是时变的随机的, 或者是遵循某一规则的混沌的。需要从不同的方向去观察数据[7] [8], 找到最能充分挖掘数据特征的最优投影方向。设  $m$  维单位向量, 即各指标的投影方向向量, 第  $i$  个样本在一维线性空间的投影特征值  $z_j$  的表达式为:

$$z_j = \sum_{j=1}^m a_j x_{ij}$$

$$\begin{cases} a_j > 0 \\ \sum_{j=1}^m a_j^2 = 1 \end{cases}$$

为构造  $m$  维单位投影向量  $a$ , 设计了随机函数 Rand\_Value, 用以产生满足上述规则的  $a_{ij}$ 。另一方面, 也是为了适应黑盒型系统。使得融合算法有更好的适应性, 同时解决了固定权值的不灵活性。增强了系统的弹性度和自适应度。

### 2.3.2. 投影方向优良评估设计

一个优良的评估值集合, 应当满足概率论高斯分布的基本规则。即优、差的评估值所占比例较小, 中等的评估值所占比例较大。因此在评估最优投影方向的规则设计上, 要求投影值  $Z_i$  的分布特征满足局部投影点尽可能密集, 凝聚成若干点团。而整体上的投影点团尽可能的散开[7]。为此, 构造投影指标函数:

$$\max Q(a) = S_a D_a$$

其中  $S_a$  为投影值  $Z_i$  的标准差,  $D_a$  为投影值  $Z_i$  的局部密度。这里选出密度窗口半径为 0.1 倍的  $S_a$ 。对  $Z_i$  内的元素进行半径穷举窗口比较, 半径小于窗口值  $0.1S_a$  则通过, 送入评估值累加器, 否则不进行累加。最终得到的值即是对一个投影向量优良评估的值[9]。采用的思想为聚类 K-MEANS [10], 鉴于实际的系统状态总是在局部稳定和全区分散的实际情况, 投影方向的评估规则也是根据是否符合高斯概率分布和聚类离散特征进行设计的。评估投向方向优良性的具体流程如图 4 所示。

## 2.4. 模拟退火法原理及应用

### 2.4.1. 模拟退火法的原理

模拟退火法是从物理退火中得到启发, 从而设计出的最优化算法。粒子在某个温度  $T$  时, 固体所处的状态具有一定的随机性, 而这些状态之间的转换能否实现由 Metropolis 准则决定[11]。Metropolis 准则

定义了物体在某一温度  $T$  下从状态  $i$  转移到状态  $j$  的概率  $P$ , 其公式如下:

$$P_{ij}^T = \begin{cases} 1, E(j) \leq E(i) \\ e^{-\left(\frac{E(j)-E(i)}{KT}\right)}, \text{其他} \end{cases}$$

上式中  $e$  为自然对数,  $E(i)$  和  $E(j)$  分别表示物体在状态  $i$  与状态  $j$  下的内能,  $E(j) - E(i)$  表示内能的增量,  $K$  是玻尔兹曼常数。

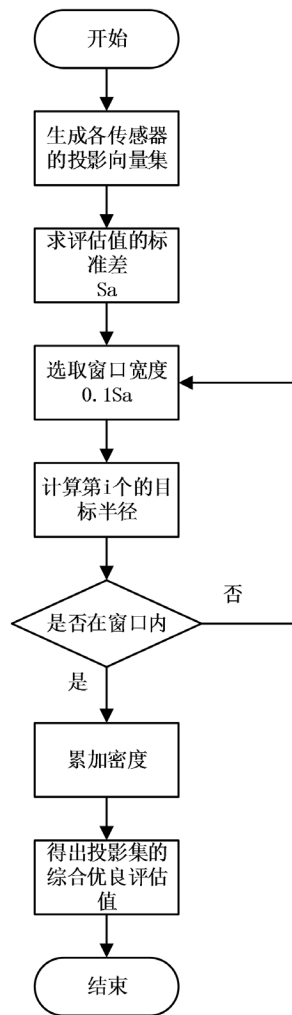


Figure 4. Projection direction good evaluation flowchart  
图4. 投影方向优良评估流程图

#### 2.4.2. 模拟退火法在本文中的应用

本文对模拟退火法的应用是对投影向量  $w_m$  按照一定的准则生成一个评判  $w_m$  优良的标量  $Tg$ , 进行模拟退火, 与新的  $w'_m$  生成的标量  $Tg'$  进行比较。如果  $Tg' > Tg$ , 则接受新的变化  $w'_m$ 。否则如果  $e^{-\left(\frac{Tg'-Tg}{KT}\right)} > rand()$  (这里的  $rand()$  指的是 0~1 内的随机数), 那么也接受新的变化。否则不会接受新的变化。在冷却过程中, 冷却反弹的几率, 即在  $Tg' < Tg$  情况下仍然接受新变化的几率, 会随着温度的下降逐渐下降, 即温度越低就越稳定的退火原理。所以迭代过程中概率反弹会越来越小, 这也符合自然降温

或优胜劣汰的生存规律。最终, 冷却结束, 保留的即是最优的投影方向。选取的测试集迭代大约 45,900 次时得到最优解。

### 2.4.3. 模拟退火法流程

根据以上原理进行退火算法设计, 流程如图 5 所示。

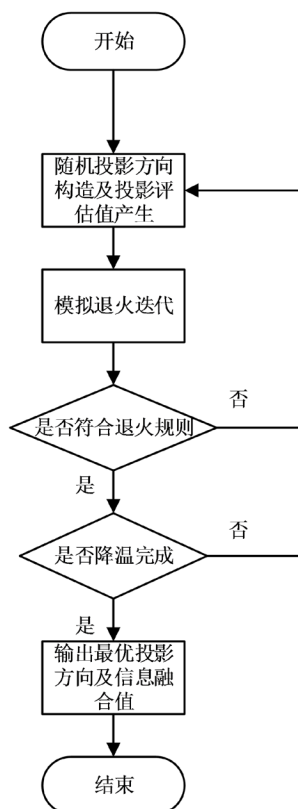


Figure 5. Simulated annealing process

图5. 模拟退火法流程

## 3. 结果

本文采取的算法测试集为 Kaggle 数据库中一农业养殖大棚中 6 种传感器所测得的 100 组样本数据(土壤含水率传感器、温度传感器、湿度传感器、光照强度传感器、二氧化碳传感器、土壤 pH 值传感器)部分样本数据如表 1 所示。

Table 1. Partial sensor data

表 1. 部分传感器数据

土壤含水率(g/cm <sup>3</sup> )	二氧化碳浓度(ppm)	光照强度(lx)	最适温度偏差(°C)	最适湿度偏差(rh)	土壤相对最适 pH 值偏差
1.052	975	139.45	4.366	19.4	0.928
1.011	761	82.00	0.981	24.6	1.617
1.176	690	119.28	2.891	4.4	2.981
1.272	659	185.45	2.257	24.1	2.409
1.379	685	146.09	0.558	19.4	2.311

## Continued

1.354	554	157.12	6.119	14.9	0.406
1.181	779	82.97	7.877	7.6	0.850
1.425	826	195.98	9.534	25.5	1.583
1.409	806	69.09	8.707	24.6	1.646
1.185	904	87.21	2.370	38.3	2.530
1.357	515	158.66	9.946	26.0	2.933
0.910	308	186.03	4.661	26.2	0.416
0.992	392	159.10	9.795	6.4	0.922
1.038	699	159.88	7.311	33.8	1.026
0.954	959	60.95	7.040	6.2	2.121
1.152	917	113.80	6.577	19.9	1.420
1.167	967	153.03	4.767	8.9	1.221
1.142	962	173.69	3.890	8.9	0.896
1.038	888	144.69	9.289	5.4	1.306
1.422	766	130.20	10.132	33.0	0.215
1.074	845	77.44	4.975	6.3	2.341
1.091	420	59.48	7.779	32.7	2.951
1.186	594	88.03	2.135	27.2	2.516
1.214	500	158.03	6.294	32.1	1.685
1.305	565	188.81	4.137	5.6	2.356
1.362	928	72.77	6.477	13.8	0.519
1.375	506	92.01	4.656	19.0	1.781
0.902	426	185.62	4.568	2.7	2.634

经过信息融合后得到的样本综合评估值如图 6 所示, 100 组样本产生的多传感器信息融合的综合评估值随机分布, 而且中部的评估值较为集中。对得到的投影向量进行散点图分析, 该数据的分布满足 K-MEANS 聚类 and 概率论的原则。多传感器信息融合得到的综合指标散点分布图如图 7 所示, 可以明显的发现局部的点聚集成团, 而整体又尽可能的散开。

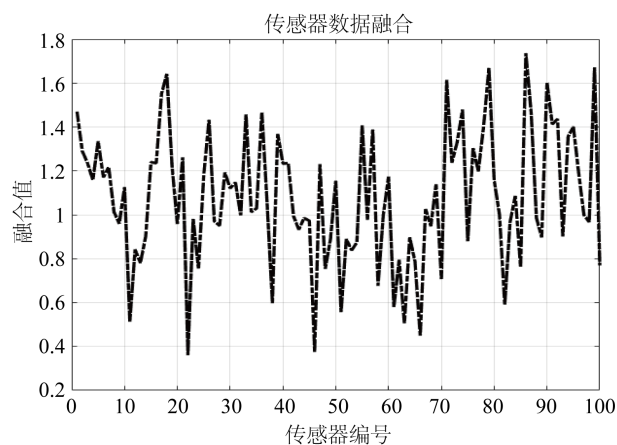


Figure 6. Sample comprehensive evaluation value  
图6. 样本综合评估值



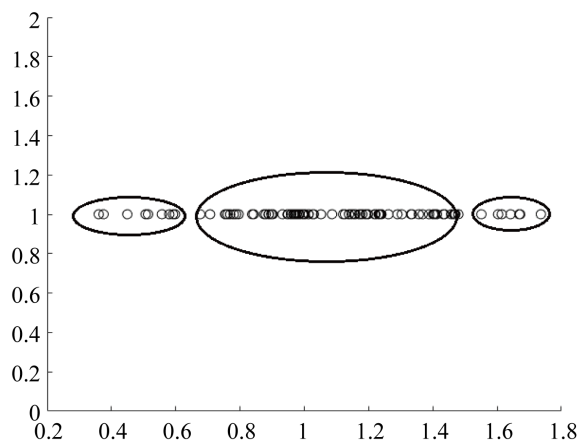


Figure 7. Sample comprehensive evaluation scatter distribution  
图7. 样本综合评估散点分布图

回归到原始的表格数据, 这里我们观察评分最高点样本、最低点样本、平均点样本。图中最高点样本、最低点样本、平均点样本编号为 86、22、42。

86 号样本数据:

1.340	894	151.55	0.598	4.1	1.825
-------	-----	--------	-------	-----	-------

22 号样本数据:

1.091	420	59.48	7.779	32.7	2.951
-------	-----	-------	-------	------	-------

42 号样本数据:

1.160	995	129.37	8.672	39.5	1.365
-------	-----	--------	-------	------	-------

数据从左至右分别为土壤含水率、二氧化碳浓度、光照强度、最适温度偏差、最适湿度偏差、最适土壤 pH 值偏差。前三项为正向指标, 而后三项为负相关指标。经过信息融合评分值最高的样本的土壤含水率、二氧化碳浓度、光照强度均较高, 最适温度偏差、最适湿度偏差、最适土壤 pH 值偏差均较低。虽然在 pH 值数据上低于平均的水平, 但是考虑总体上的信息融合是综合各项指标, 对于所得出的结果仍是可信的。

对于非融合的单—传感器数据, 传统的做法是将各项数据逐一传送进行各类数据的阈值判断。传统的非融合单—传感器数据采集与决策模式, 如图 8 所示。

不考虑传感器故障的情况下, 分布式的单—传感器网络, 在不进行信息融合的情况下, 很难对于系统有一个全局性的把握。特别是对于复杂、庞大的传感器系统, 单一的逐次处理可能会造成控制时机的丧失和资源的浪费以及数据的冗余。下面以一些数据图表进行简要的说明:

图 9 中, 数据庞多且种类很多, 还可能包含正负指标不一样的数据类型。如果仍使用传统的非融合单—传感器信息技术处理就很难得到系统整体的情况, 大批数据的逐次处理对于系统资源的浪费和对后续系统的扩展不利。而采用多传感器信息融合算法所得到的结果更加清晰简明, 有利于决策。

对于非融合单—传感器, 假设初始系统的处理复杂度为  $O_{(n)}$ , 规定每次的系统更新扩展会使得系统在原来的基础上增大为 1.13 倍的复杂度。第一次更新后系统的复杂度为  $O_{(1.13n)}$ , 第二次更新后系统的复杂度

为  $O_{(1.2769n)}$ , 第三次……计算多次发现系统将会承受几十倍原来复杂度。现假设使用多传感器信息融合技术的系统初始复杂度为  $O_{(\frac{n}{2}+0.001)}$ , 第  $k$  次的复杂度为  $O_{(\frac{1.13^k}{2}+0.001)*n}$ 。传感器数据融合技术将复杂度为  $O_{(n)}$  的系统降低至  $1/2$ , 同时增添  $0.001$  倍的  $O_{(n)}$  总体数据指标处理复杂度。得到的系统复杂程度发展特性图如图 10 所示。从图 10 可看出经过融合后的计算复杂度在系统不断发展后, 所减少的复杂度由 3% 升高至 15%。

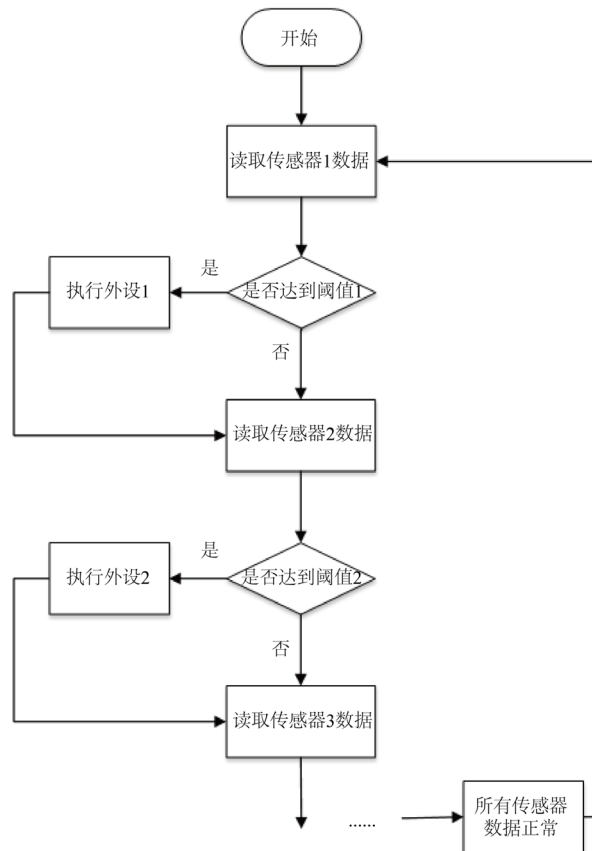


Figure 8. Traditional non-fusion single sensor data acquisition and decision-making mode

图8. 传统的非融合单一传感器数据采集与决策模式

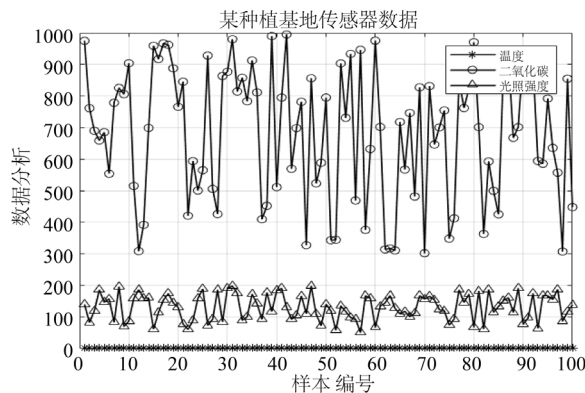
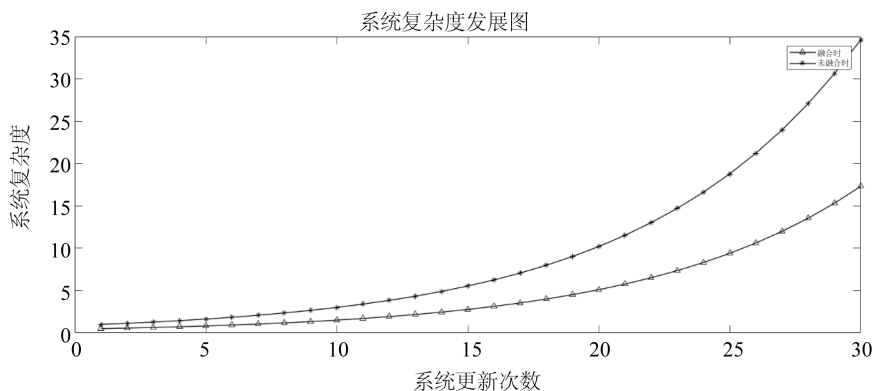


Figure 9. Data graph of multiple sensors

图9. 多个传感器的数据图表



**Figure 10.** The development chart of system complexity using fusion technology and non-fusion technology

**图10.** 采用融合技术与非融合技术系统复杂度发展对比图

在本文所选取的测试集指标为 6 个, 组数 100, 前 3 项指标为正性指标, 后 3 项为负性指标。算法给出的评估值符合正态分布, 样本评估值分布如图 7 所示。经过回归元数据表中查对正负峰值(样本 86、22)、平均值(样本 42)。发现峰值综合正向指标和为 1047 接近各样本正性数据抽取最大值之和 1194.568; 峰值综合负向指标和为-6.55, 接近各样本负性数据抽取最大值之和-1.955。综合评分最低的样本正性指标和为 481.054, 接近各样本正性数据抽取最小值之和 352.857; 负性指标和为-43.46, 接近各样本负性数据抽取最小值之和-52.641。另外, 通过系统聚类窗口半径的调整可以改变系统状态的分层等级; 通过改变模拟退火迭代器的迭代次数可以在精准度和实时性之间进行有效的平衡; 通过改变反弹机会的窗口步距可以改变系统自然状态奇点对于评估值的影响。综上, 本设计可以给多传感器信息融合系统更大的灵活性、适应性。

#### 4. 结论与讨论

本文利用投影寻踪法和模拟退火法结合, 吸收了加权平均算法的权值判别、概率分布型算法的概论权值和随机概论向量、神经网络渐缩迭代方法及 Sigmoid 数据统一方法, 设计了一种基于加权平均法的多传感器信息融合算法。在加权算法的基础上进行了优化扩展。经过样本测试集的测试, 发现算法的计算复杂度较单一传感器信息算法提升了 41.6%左右, 测试结果的峰值与测试集个样本抽取最大最小峰值基本吻合, 有着较优异评估的结果。该算法模型适用范围广, 对于正负性指标杂糅和黑盒型系统有专门的处理, 结构简洁灵活, 可扩展性强。但有的地方还需进一步的改进, 比如在算法运算速度和与经验知识结合处理方面, 有待提升。

5G 时代, 万物互联。多传感器信息融合技术给空间传感器网络带来智能化的监控管理。对于多传感器信息融合技术的算法研究将持续进行。随着智能传感技术的不断发展和融合技术的不断完善, 人们对于系统或者设备的整体状态的把握将更加精确。智能生活, 万物互联的未来已然来临。

#### 基金项目

国家自然科学基金(11703009) (针对平移误差检测的电容式边缘传感器改进研究)。

#### 参考文献

- [1] 赵宗贵. 信息融合技术现状、概念与结构模型[J]. 中国电子科学研究院学报, 2006(4): 305-312.
- [2] Zhou, Y.H., Zuo, Z.T. and Xu, F.R. (2019) Origin Identification of Panax Notoginseng by Multi-Sensor Information

Fusion Strategy of Infrared Spectra Combined with Random Forest. *Spectrochimica Acta-Part A: Molecular and Biomolecular Spectroscopy*, Vol. 226, 5 February 2019. <https://doi.org/10.1016/j.saa.2019.117619>

- [3] 刘国凯. 基于多传感信息融合的旋转机械故障诊断研究[D]: [硕士学位论文]. 贵阳: 贵州大学, 2018.
- [4] 赵毅. 数字滤波的算术平均法和加权平均法[J]. 仪表技术, 2001(4): 41-44.
- [5] 周治利. 基于概率假设密度的多传感器多目标跟踪技术研究[D]: [硕士学位论文]. 杭州: 杭州电子科技大学, 2017.
- [6] Šárka, B., Peter, F., Thomas, O., Christian, B. and Maia, R. (2019) Robust and Sparse K-Means Clustering for High Dimensional Data.
- [7] 张欣莉, 丁晶, 李祚泳, 金菊良. 投影寻踪新算法在水质评价模型中的应用[J]. 中国环境科学, 2000(2): 187-189.
- [8] 王顺久, 张欣莉, 丁晶, 侯玉. 投影寻踪聚类模型及其应用[J]. 长江科学院院报, 2002(6): 53-55+61.
- [9] 金菊良, 魏一鸣, 付强, 丁晶. 农业生产力综合评价的投影寻踪模型[J]. 农业系统科学与综合研究, 2001(4): 241-243.
- [10] 任恒妮. 大数据 K-means 聚类算法的研究与应用[J]. 信息技术, 2019, 43(11): 20-23.
- [11] 饶太春, 兰林强, 罗伟林. 模拟退火法在协同优化中的应用[J]. 福州大学学报(自然科学版), 2018, 46(3): 379-385.