

基于 D-S 证据理论的智能温室环境控制决策融合方法

孙力帆¹ 张雅媛¹ 郑国强¹ 冀保峰¹ 何子述²

(1. 河南科技大学信息工程学院, 洛阳 471023; 2. 电子科技大学电子工程学院, 成都 611731)

摘要: 在无线传感器网络下的智能温室环境控制系统中,农作物的生长通常受多种环境因子共同作用。根据温室环境控制系统的实际需求建立基于 Dempster-Shafer(D-S)证据理论的决策框架,并提出了一种数据预处理和决策融合方法。首先,使用箱线图检测测量数据中的异常值,考虑到现有直接剔除异常数据处理方法的弊端,提出了一种异常数据自适应修正方法;然后,利用加权平均距离聚类处理更新后的数据;最后,根据所提出的基于加权相似度的基本概率分配方法结合 D-S 证据理论进行融合,为温室环境控制做出正确决策。实验结果表明,箱线图检测异常数据更为准确,其检测率比狄克逊准则高近 19.2%,对于不确定性融合结果,本文提出的基于加权相似度的基本概率分配方法相比现有方法降低了 1~2 个数量级,不仅可以提高温室环境参数融合精度,加快收敛速度,同时还能有效地降低决策风险。

关键词: 温室环境控制; 决策融合; D-S 证据理论; 异常值自适应修正; 聚类分析

中图分类号: S625.5; TP274+.2 **文献标识码:** A **文章编号:** 1000-1298(2018)01-0268-08

Approach to Decision Fusion for Intelligent Greenhouse Environmental Control Based on D-S Evidence Theory

SUN Lifan¹ ZHANG Yayuan¹ ZHENG Guoqiang¹ JI Baofeng¹ HE Zishu²

(1. School of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

2. School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: In the intelligent greenhouse environment control system using wireless sensor networks, the crops growing is usually affected by various factors from the environment. The greenhouse control system makes inappropriate decisions based on the received measurements of wireless sensor networks, which will lead to go against the growth of crops. In view of this, a novel decision-making framework based on Dempster-Shafer(D-S) evidence theory was established to meet practical requirements of the greenhouse environment control system. Moreover, two approaches of the data preprocessing and the decision fusion were proposed, respectively. Firstly, the measuring outliers data were detected by using the box-plot, and then an effective approach was proposed to correct them adaptively, which overcame the disadvantage of directly removing the outlier data in existing approaches. The corrected measuring data were also clustered by using the weighted average distance. Finally, a novel basic probability assignment approach was proposed to make correct decisions for the control of greenhouse environment based on the D-S evidence theory. Experimental results demonstrated that the outliers detection rate of the box-plot was more accurate than that of the Dixon criterion (nearly 19.2%). Compared with existing approaches, the fusion performance of the uncertainty was reduced by 1~2 order-of-magnitude in the proposed basic probability assignment approach, which not only increased precision of environmental indicators for the greenhouse control, but also accelerated the convergence process and reduced the risk of decision-makings effectively.

Key words: greenhouse environment control; decision fusion; D-S evidence theory; adaptive outlier correction; cluster analysis

收稿日期: 2017-08-26 修回日期: 2017-11-10

基金项目: 国家自然科学基金项目(U1504619、U1404615、61671139)、教育部产学研合作协同育人项目(201602011005)和河南省科技攻关计划项目(162102210073)

作者简介: 孙力帆(1982—),男,讲师,博士,主要从事多传感器信息融合理论与技术、智能感知与信息处理研究,E-mail: lifan_sun@126.com

0 引言

现代智能温室环境控制系统受到越来越多的关注,与此相关的无线传感器网络信息融合技术更是得到了广泛应用^[1-5]。无线传感器网络下智能温室环境控制的核心是智慧管理和智能决策,即使用大量密集部署在温室监控区域的智能传感器节点对温室大棚内温湿度、光和有效辐射、二氧化碳、光照等小气候环境参数数据进行自动采集和分析,以智能调控温室达到适宜植物生长的范围。但是,由于受到外部环境和无线传感器内部构造等多种因素的影响,传感器测得的数据往往具有一定不确定性,而 Dempster-Shafer (D-S) 证据理论能够对其进行有效处理,且具有很强的灵活性和可靠性,已被许多专家系统所应用。尽管证据理论有许多优点,但当证据之间存在高度冲突时,D-S 证据理论的处理结果往往与常理相悖。目前,解决证据冲突问题主要有以下 2 类方法:①在不改变融合规则的前提下,对传感器量测数据进行预处理,即采用剔除异常数据的方法或聚类分析,以减少数据波动,从而使得数据在决策融合之前达到理想状态降低证据冲突程度,提高融合精度^[6-12]。虽然能从不同程度上减少证据冲突带来的数据误差,降低数据间波动对融合结果的影响,但是它们几乎都是剔除掉原始数据中存在的异常值,必然导致样本数据减少,不利于提高后续的整体融合精度。②通过改进证据理论融合规则以达到融合后置信函数的合理分配,或在融合之前选取各个证据最优的置信函数^[13-22]。当证据间高度冲突或者完全冲突时,虽然修改证据理论融合规则可以使置信函数合理分配,但是这种做法有可能破坏证据融合中的交换律、结合律等优良特性,不能真正达到降低冲突程度的目的。

上述两类方法在处理决策融合时均具有一定局限性。本文结合它们各自优点提出一种温室环境控制决策融合框架。首先,使用箱线图检测出原始量测数据中存在的异常值;然后对其进行预处理,提出一种异常数据自适应修正算法以减少误差,并对修正更新完成后的数据进行系统聚类;最后,根据本文提出基于加权相似度的基本概率分配 (Basic probability assignment, BPA) 方法结合 D-S 证据理论进行数据融合。

1 数据预处理

1.1 异常数据自适应修正算法

一般地,传感器测量数据往往都是随机的且有

可能存在异常值,是指测量数据集合中严重偏离大部分数据所呈现趋势的小部分数据点,由于传感器自身或数据传输的原因,使得测量数据序列中存在某些错误的测量。数据中存在的异常值会使后续的决策融合结果严重偏离预想状态,继而产生较大误差。常用检测异常值的方法有 3 σ 法则、狄克逊准则和 z 分数方法等。以上方法均是假定测量数据呈正态分布,而实际上它们的分布往往是未知的,此外,这些方法判断异常值的标准是通过计算批量数据的均值和标准差来实现,但是以上 2 个统计量易受异常值约束,使得异常数据检测率偏低。本文采用箱线图检测并对其自适应修正。

定义 1:箱线图是用来体现数据分散情况的统计图,可以表示数据的分布差异,它由 5 个数据统计量组成:最小值 $X_{\min}$ 、下四分位数 Q_1 、中位数 Q_2 、上四分位数 Q_3 、最大值 $X_{\max}$ 。箱线图结构如图 1 所示,红色圆圈为异常值标记。

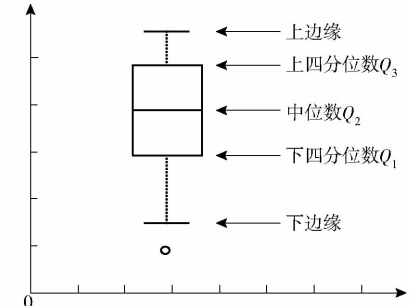


图 1 箱线图结构

Fig. 1 Structure of box-plot

待统计数据按从小到大排列 $X: (x_1, x_2, \dots, x_N)$, 中位数 Q_2 定义为第 50% 数据, 下四分位数 Q_1 定义为第 25% 数据也可定义为 $[X_{\min}, Q_2]$ 的中位数, 上四分位数 Q_3 定义为第 75% 数据同理可定义为 $[Q_2, X_{\max}]$ 的分位数^[9]。以 6 个数据为例, 按其从小到大顺序排列为 $x_1, x_2, \dots, x_6$, 具体计算方法为:

Q_1 的位置

$$\begin{cases} n_1 = \frac{6+1}{4} = 1.75 \\ Q_1 = x_1 + (x_2 - x_1)(n_1 - \lfloor n_1 \rfloor) \end{cases} \quad (1)$$

Q_2 的位置

$$\begin{cases} n_2 = \frac{6+1}{2} = 3.5 \\ Q_2 = x_3 + (x_4 - x_3)(n_2 - \lfloor n_2 \rfloor) \end{cases} \quad (2)$$

Q_3 的位置

$$\begin{cases} n_3 = \frac{3(6+1)}{4} = 5.25 \\ Q_3 = x_5 + (x_6 - x_5)(n_3 - \lfloor n_3 \rfloor) \end{cases} \quad (3)$$

异常点的上下边缘为

$$\begin{cases} F_U = Q_3 + \frac{\beta}{2} I_{QR} \\ F_L = Q_1 - \frac{\beta}{2} I_{QR} \end{cases} \quad (4)$$

其中 $I_{QR} = Q_3 - Q_2$
式中 I_Q —— Q_1 与 Q_3 之间的距离
 F_U ——异常值上边缘分界点
 F_L ——异常值下边缘分界点

显然,使用箱线图对数据分布进行描述,异常值清晰可见。其判断标准是大于 $Q_3 + 1.5I_{QR}$ 或者小于 $Q_1 - 1.5I_{QR}$ (即式(4)中 $\beta = 3$),经验表明该标准的选取在处理批量数据时有较高的鲁棒性。相较于现有主流的 3δ 法则、狄克逊准则等异常值检测方法,箱线图的优势在于它从实际数据出发,无需假设其统计分布,并且四分位数表示形式不易受异常数据影响。

针对异常数据的处理通常采用直接将其剔除的办法,但是仅仅剔除异常值不但会使样本数量减少,而且有可能使得每组数据的维数不尽相同,从而导致精度降低。因此,本文提出了一种基于专家系统的异常数据自适应修正算法,能够在不改变样本数量的前提下通过对其进行修正,提高批量数据的鲁棒性。

在文献[9]的温室控制专家系统表中,4种决策分别对应的环境参数为 25.3、31.3、39.3、35.4℃,则所提出的算法根据以上均值进行如下修正。

情况 1: $x_1 > x_m, \Delta = x_1 - x_m$

$$\hat{x} = x_1 - \frac{\Delta}{2} \quad (\Delta \leq 5) \quad (5)$$

$$\begin{cases} x' = x_1 - \frac{\Delta}{2} \\ \hat{x} = x' - \frac{|x' - x_m|}{2} \end{cases} \quad (5 < \Delta \leq 10) \quad (6)$$

$$\begin{cases} x' = x_1 - \frac{\Delta}{2} \\ x'' = x' - \frac{|x' - x_m|}{2} \\ \hat{x} = x'' - \frac{|x'' - x_m|}{2} \end{cases} \quad (10 < \Delta \leq 15) \quad (7)$$

情况 2: $x_1 < x_m, \Delta = x_m - x_1$

$$\hat{x} = x_1 + \frac{\Delta}{2} \quad (\Delta \leq 5) \quad (8)$$

$$\begin{cases} x' = x_1 + \frac{\Delta}{2} \\ \hat{x} = x' + \frac{|x_m - x'|}{2} \end{cases} \quad (5 < \Delta \leq 10) \quad (9)$$

$$\begin{cases} x' = x_1 + \frac{\Delta}{2} \\ x'' = x' + \frac{|x_m - x'|}{2} \\ \hat{x} = x'' + \frac{|x_m - x'|}{2} \end{cases} \quad (10 < \Delta \leq 15) \quad (10)$$

式中 x_1 ——异常值 x_m ——专家系统值
 $\hat{x}$ ——修正值

一般地,温室大棚中的小气候参数,如温度、光照强度和二氧化碳体积分数的原始量测数据(含异常值)范围是在专家系统均值上下波动不超过 15 (不包括量纲),只是温度和光照强度波动范围会比二氧化碳体积分数小一些。文中分 3 段依次对其进行修正,其最终目的是让异常值逐渐收敛于专家系统值。

情况 1:异常值 x_1 大于专家系统均值 x_m 时,其差值记为 $\Delta, \hat{x}$ 为修正后值,当 $\Delta \leq 5$ 时,则 $\hat{x} = x_1 - \Delta/2$,称为 1 次收敛运算;当 $5 < \Delta \leq 10$,对异常值进行 2 次收敛运算;当 $10 < \Delta \leq 15$,则对异常值进行 3 次收敛运算。

情况 2:异常值 x_1 小于专家系统均值 x_m 时,其差值记为 $\Delta, \hat{x}$ 为修正后的值,当 $\Delta \leq 5$ 时, $\hat{x} = x_1 + \Delta/2$,同样称为一次收敛运算;当 $5 < \Delta \leq 10$ 与 $10 < \Delta \leq 15$ 时,与情况 1 一样分别对异常值进行 2 次、3 次收敛运算。当所有异常值修正完成后,通过箱线图观察更新数据是否还有异常值,如果有,继续对其进行修正直至没有异常值。本文研究表明,完成全部数据更新一般不超过 3 次迭代修正。

1.2 不同聚类准则下的数据聚类融合

当所有数据更新完成后,需要对个数较多且分散的数据聚类融合分析使之系统化。本文首先利用凝聚的层次聚类对数据处理,把每个数据作为一类,计算数据距离并根据类间距离进行合并直至所有数据都归为一类。

一般地,使用不同的距离准则会产生不同的聚类结果^[9]。充分考虑不同环境参数下测量数据(如温度(℃)、光照度(klx)、二氧化碳体积分数(μL/L))对决策结果的影响,选择马氏距离计算数据间的距离,以提高聚类精度。设 n 维列向量 $X = (x_1, x_2, \dots, x_n)^T$,均值为 $U = (u_1, u_2, \dots, u_n)^T$, u_i 表示第 i 个变量的期望值即 $u_i = E(x_i)$,协方差矩阵为 Σ ,其马氏距离定义为 $D_n(X) = \sqrt{(X - U)^T \Sigma^{-1} (X - U)}$,其中协方差矩阵为

$$\Sigma = E[(X - U)(X - U)^T] = \begin{bmatrix} E[(x_1 - u_1)(x_1 - u_1)] & E[(x_1 - u_1)(x_2 - u_2)] & \cdots & E[(x_1 - u_1)(x_n - u_n)] \\ E[(x_2 - u_2)(x_1 - u_1)] & E[(x_2 - u_2)(x_2 - u_2)] & \cdots & E[(x_2 - u_2)(x_n - u_n)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(x_n - u_n)(x_1 - u_1)] & E[(x_n - u_n)(x_2 - u_2)] & \cdots & E[(x_n - u_n)(x_n - u_n)] \end{bmatrix} \quad (11)$$

由式(11)可得马氏距离矩阵

$$D_n(X) = \begin{bmatrix} 1 & d_{12} & \cdots & d_{1n} \\ d_{21} & 1 & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & 1 \end{bmatrix} \quad (12)$$

根据式(12)中 $D_n(X)$ 选取不同的距离准则对修正后的数据进行聚类:一是使用最小距离聚类准则首先在 $D_n(X)$ 中查找最小值,然后将距离最小的这 2 个数据进行聚类得到新的一类值,再重新计算距离矩阵,依次聚类得到最终的聚类值。二是使用平均加权距离聚类准则对分布密集的数据进行聚类,继而最终将其归为一类。研究表明,2 种距离准则下的聚类结果基本一致,但使用加权平均距离可得到比最小距离更高的聚类精度,且更接近实际。

2 基于 D-S 证据理论的温室环境控制决策融合

温室环境控制系统的目的是根据农作物实际生长需求做出某种控制决策,但直接使用上述聚类融合处理后的数据仍然会因其本身所具有的不确定性而产生一定误差,以致无法精确调控温室达到适宜农作物生长的范围。本文引入 D-S 证据理论来解决数据不确定性所产生的误差而带来的决策风险问题。

2.1 经典证据理论

D-S 证据理论是将 2 个或多个基本概率分配函数通过理论规则融合成新的概率分配函数,并将其作为最终的决策依据,用以提高整个决策可靠性。

(1) 基本概率分配函数(BPA)

在温室环境控制系统中,假设所有可能的决策结果组成一个完备集合 Θ ,其中各个元素相互独立。那么, Θ 中各焦元的 m 函数分配定义为

$$m: 2^\Theta \rightarrow [0, 1]$$

满足

$$m(\Phi) = 0 \text{ 且 } \sum_{A \subseteq \Theta} m(A_i) = 1 \quad (13)$$

其中满足 $m(A_i) > 0$ 的元素 A_i 称为焦元, $A_i (i = 1, 2, \dots, n) \in \Theta$ 。

(2) m 函数合成规则

$\forall A_i, A_i (i = 1, 2, \dots, n) \in \Theta$ 均不为空集, Θ 上 n 个 m 函数 $m_1, m_2, \dots, m_n$ 的合成规则为

$$\begin{cases} m(A) = \frac{1}{K} \sum_{A_1 \cap A_2 \cap \cdots \cap A_n = A} m_1(A_1) m_2(A_2) \cdots m_n(A_n) \\ K = \sum_{A_1 \cap A_2 \cap \cdots \cap A_n \neq \Phi} m_1(A_1) m_2(A_2) \cdots m_n(A_n) \end{cases} \quad (14)$$

2.2 概率分配算法

当证据之间存在高度冲突时,直接使用上述经典证据理论有可能会偏离实际结果而做出错误判断。本文提出了一种基于加权相似度的概率分配算法。

在 1.2 节中得到的聚类值集合 X 表示为 $(x_1, x_2, \dots, x_n)$, 专家系统均值集合 ω 表示为 $(\omega_1, \omega_2, \dots, \omega_n)$, 辨别框架 Θ 表示为 (L_1, L_2, L_3, L_4) , 其中 $L_i (i = 1, 2, 3, 4)$ 代表最终的 4 个决策内容。以数据 x_1 为例进行分析,首先根据聚类值与系统均值之间的距离得出相似度,相似度越高说明此数据的信任度越高、概率分配也越高,然后根据加权系数进行调整。令 $\omega_1 \leq x_1 \leq \omega_2$, ω_1 和 ω_2 分别对应的决策为 L_1 和 L_2 , 分别计算 x_1 与 ω_1, ω_2 的距离和相似度为

$$\begin{cases} d_1 = \frac{|x_1 - \omega_1|}{x_1 + \omega_1} \\ d_2 = \frac{|x_1 - \omega_2|}{x_1 + \omega_2} \end{cases} \quad (15)$$

$$\begin{cases} s_1 = 1 - d_1 \\ s_2 = 1 - d_2 \end{cases} \quad (16)$$

式中 d_1 —— x_1 与 ω_1 的距离

d_2 —— x_1 与 ω_2 的距离

s_1 —— x_1 与 ω_1 的相似度

s_2 —— x_1 与 ω_2 的相似度

由式(15)、(16)可知距离越小,两者的相似度越大。如果 $d_1 > d_2$, 则表明 x_1 相较于 ω_1 更接近 ω_2 , 对决策 L_2 也更信任,则所提出的基本概率分配函数为

$$\begin{cases} m(L_1) = (1 - d_1) e^{-\frac{|x_1 - \omega_1|}{\omega_2 - \omega_1} p_1} \\ m(L_2) = (1 - d_2) e^{-\frac{|x_1 - \omega_2|}{\omega_2 - \omega_1} p_2} \\ m(L_3) = 0 \\ m(L_4) = 0 \\ m(\Theta) = 1 - m(L_1) - m(L_2) - m(L_3) - m(L_4) \end{cases} \quad (17)$$

式中 $e^{-\frac{|x_1 - \omega_1|}{\omega_2 - \omega_1} \rho_1}$ —— L_1 概率分配加权因子

$e^{-\frac{|x_1 - \omega_2|}{\omega_2 - \omega_1} \rho_2}$ —— L_2 概率分配加权因子

参数 ρ_1 和 ρ_2 的设定是为了避免信任度过大而偏离实际结果,实验表明 $\rho_1 = 3$ 、 $\rho_2 = 5$ 时结果最为精确。

3 实验结果与性能对比

实验采用文献[9]中提到的温度(℃)、光照度(klx)、二氧化碳体积分数(μL/L)这3个环境参数,并选取早中晚6个时刻对以上3个参数进行数据采集。每个时刻的每个参数各采集26个数据,记为 $\mathbf{X}:(x_1, x_2, \cdots, x_{26})$,决策结果共分为4个:开启卷膜、开启卷膜和启动风机、启动风机与湿帘、无动作,即 $\mathbf{D}:(L_1, L_2, L_3, L_4)$ 。最后,按照本文提出的方法在每个时刻对各个环境参数下的测量数据进行聚类,并根据同一时刻的3个环境参数做出决策。

3.1 传感器异常数据检测及其修正

以传感器采集到的温度数据为例,图2和图3分别是6个时刻原始数据和修正后数据的箱线图表示。由图3可以看出,箱线图的箱体宽度不一,中位数的位置也不同,这取决于测量数据的分散程度和传感器的精确度,但是数据整体呈正态分布,中午温度最高,早晨和傍晚温度较低。显然,图中每个时刻原始温度数据都存在异常值,而修正之后数据如图3所示,可见修正后异常数据不复存在,且没有改变原始数据的统计分布。

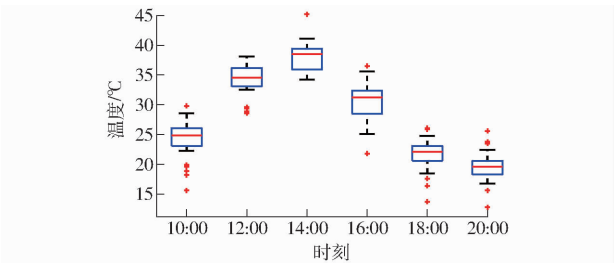


图2 原始温度数据分布图

Fig. 2 Distribution of raw temperature data

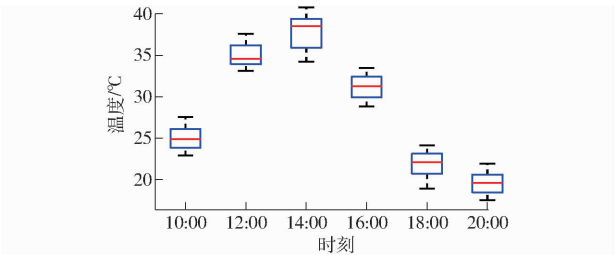


图3 修正后的温度数据分布图

Fig. 3 Distribution of corrected temperature data

接下来,本文对箱线图和狄克逊准则的异常数据检测率进行对比,如图4所示,由图4可见,箱线

图的异常数据检测率相比于狄克逊准则提高了约19.2%,该方法更为有效。

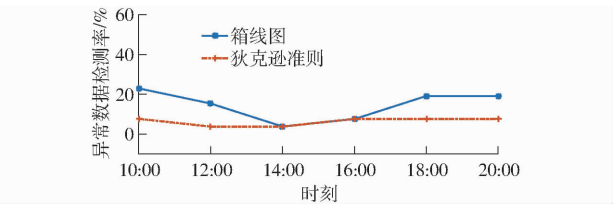


图4 异常数据检测率

Fig. 4 Detection rate of outlier data

3.2 异常数据剔除、修正后的聚类结果对比

在加权平均距离准则和最小距离准则下对修正异常数据后与直接剔除异常数据后的聚类结果进行对比(图5和图6)。特别在无线传感器网络下的智能温室环境控制系统中,由于需要检测和收集大量的传感器量测数据,需要首先着重考虑异常数据的处理工作是否会影响温室环境控制系统的稳定性,因为它是评估系统性能的根本标准。相应地,稳定性就作为一个关键评价指标被用来判定不同方法的性能。虽然直接剔除法减少了异常数据对环境控制决策所产生的不利影响,但是这种做法会造成后续聚类结果上下波动,不仅有可能导致错误的控制决策产生,而且更关键是增加了环境控制系统不稳定的风险。由图5、6可以得到,直接剔除异常值后的聚类值虽然接近专家系统值(因其直接剔除掉异常数据这一不利因素),但是相对于专家系统值波动过大,极其不稳定。而修正后的聚类值则完全根据专家系统值的变化而变化,也就是它们的聚类趋向一致,性能也更为稳定。实际的智能温室控制系统中分布着大量的传感器,则使用剔除法势必会带来更大的聚类性能波动,从而严重影响控制系统的稳定性,因此本文提出的自适应修正的方法更为稳健和优异。

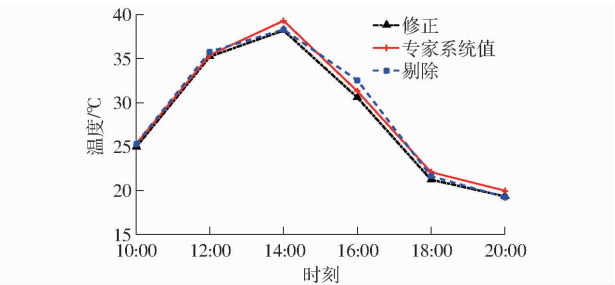


图5 加权平均距离准则下异常数据剔除、修正后的聚类结果对比

Fig. 5 Clustering comparison after removing or correcting outliers using weighted average distance

3.3 异常数据修正后不同距离准则下的聚类性能评估

此外,还对异常数据修正后2种不同距离准则

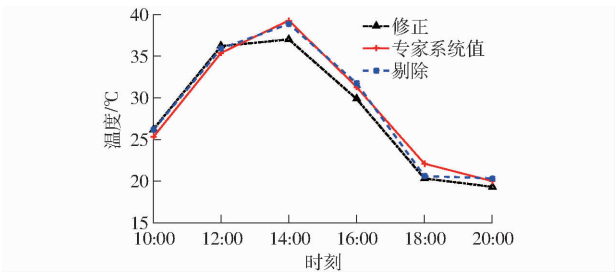


图6 最小距离准则下异常值剔除、修正后的聚类结果对比

Fig. 6 Clustering comparison after removing or correcting outliers using minimum distance

下的聚类结果进行对比(图7),由图7可得,使用平均加权距离更接近专家系统值。这是因为异常值修正后的大部分数据都与专家系统值相差无几且分布较为密集,在平均加权距离准则下对分布居中的数据进行聚类,能够使其信任度升高。而在最小距离准则下其聚类结果有可能受到分散数据的影响,因此需要首先对离散点进行聚类,但这样处理会加大离散点的信任度。显然,使用平均加权距离准则对异常值修正后的数据进行聚类更符合实际。

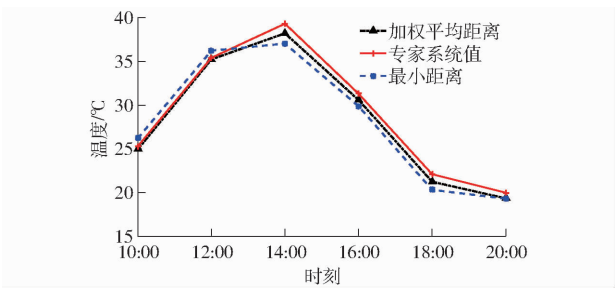


图7 两种不同距离准则下异常数据修正后的聚类结果对比

Fig. 7 Clustering comparison after correcting outliers using two different distances

修正异常值后的数据在2种距离准则下的聚类性能对比如图8所示,由图8可以看出,加权平均距离下的聚类性能较为稳定且总体高于最小距离下的聚类性能。这里特别注意的是,虽然最小距离相较于加权平距离下的聚类性能相差不大(在0.1以内),但是前者明显波动较大,其聚类结果不够可靠,显然不能被后续融合处理所采用。以上性能对比结果分析再次印证了修正数据聚类使用加权平均距离准则比最小距离准则鲁棒性更高,适用性更强。

3.4 决策融合结果对比及其性能评估

为了验证所提出基于加权相似度概率分配函数对于温室环境控制决策融合的有效性,本文与文献[9]提出的决策融合方法进行了对比分析。依次选取了4组数据进行基本概率分配实验,2种方法的概率分配分别如表1和表2所示,其决策融合对比结果见表3。从表1、2可以看出,2种方法的基本

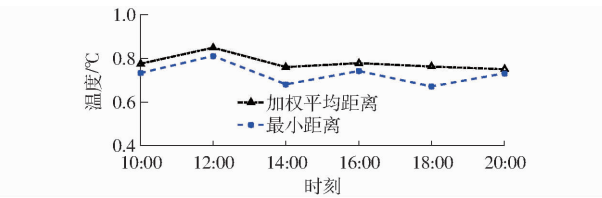


图8 两种不同距离准则下异常数据修正后的聚类性能对比

Fig. 8 Comparison of clustering performance after correcting outliers using two different distances

可信度分配效果一致,其中 t 、 c 、 i 分别代表温度、光照度、二氧化碳体积分数,但是,文献[9]提出的方法(采用均值划分)中不确定的基本概率分配全部为默认的0.1(表1)。但由于每个时刻不同温室环境参数下的数据不同,传感器精度也不同,不确定的基本概率分配也必然不同,那么显然该方法并不符合实际。

表1 基本均值划分的概率分配表

Tab. 1 Probability allocation of basic mean division

样本	证据	$m(1)$	$m(2)$	$m(3)$	$m(4)$	$m(\Theta)$
1	t	0	0	0	0.900	0.1
1	c	0.013	0	0	0.887	0.1
1	i	0	0	0	0.900	0.1
2	t	0.793	0	0	0.107	0.1
2	c	0.900	0	0	0	0.1
2	i	0.885	0	0	0.015	0.1
3	t	0.038	0.862	0	0	0.1
3	c	0	0.877	0	0.023	0.1
3	i	0	0.877	0	0.023	0.1
4	t	0	0.256	0.644	0	0.1
4	c	0	0	0.900	0	0.1
4	i	0	0.012	0.888	0	0.1

表2 基于加权相似度的概率分配表

Tab. 2 Probability allocation based on weighted similarity

样本	证据	$m(1)$	$m(2)$	$m(3)$	$m(4)$	$m(\Theta)$
1	t	0	0	0	0.983	0.017
1	c	0.048	0	0	0.920	0.032
1	i	0	0	0	0.960	0.040
2	t	0.918	0	0	0.064	0.018
2	c	0.890	0	0	0.100	0.010
2	i	0.917	0	0	0.036	0.047
3	t	0.180	0.806	0	0	0.014
3	c	0	0.847	0	0.051	0.102
3	i	0	0.876	0	0.046	0.078
4	t	0	0.041	0.861	0	0.098
4	c	0	0	0.920	0	0.080
4	i	0	0.036	0.933	0	0.031

对比如表1、表2中提出的基于加权相似度的概率分配方法,能够根据实际情况严格按照更新数据与专家系统值进行对比加权,从而大大减少了经验选取所带来的风险。此外,本文的方法相比文

献[9]中的方法不但在不确定的基本信度分配上降低了1~2个数量级,而且有效地减小了数据间的冲突程度,使得同一样本间各环境参数信任度高度匹配。

本文对两种方法中证据理论融合后的概率分配进行比较(表3),可以看出,相比于文献[9],本文的方法提高了判断精度使得决策结果更加准确。特

别注意的是,文献[9]中基于平均划分的融合结果表明其不确定性融合精度为 10^{-3} ,而从本文提出的基于加权相似度融合结果中可以观察到不确定性融合精度降为 10^{-4} 或 10^{-5} ,可见融合后的不确定性大幅度降低(即1~2个数量级),表现了极快的收敛速度。

表 3 两种方法的融合结果对比
Tab.3 Comparison of fusion results in two approaches

数据来源	证据组合	$m(1)$	$m(2)$	$m(3)$	$m(4)$	$m(\Theta)$
文献[9]	$t\oplus c\oplus i$	1.0×10^{-4}	0	0	0.998 9	1.0×10^{-3}
本文	$t\oplus c\oplus i$	8.0×10^{-5}	0	0	0.999 9	2.0×10^{-5}
文献[9]	$t\oplus c\oplus i$	0.997 3	0	0	1.6×10^{-3}	1.1×10^{-3}
本文	$t\oplus c\oplus i$	0.999 1	0	0	9.0×10^{-4}	1.0×10^{-5}
文献[9]	$t\oplus c\oplus i$	4.0×10^{-4}	0.997 9	0	6.0×10^{-4}	1.1×10^{-3}
本文	$t\oplus c\oplus i$	1.9×10^{-3}	0.997 7	0	2.0×10^{-4}	2.0×10^{-4}
文献[9]	$t\oplus c\oplus i$	0	4.0×10^{-3}	0.994 6	0	1.4×10^{-3}
本文	$t\oplus c\oplus i$	0	5.0×10^{-4}	0.999 2	0	3.0×10^{-4}

4 结 论

(1)通过采用箱线图法检测原始数据的异常值,可以达到比现有方法更高的检测率,最高可以达到22.3%。

(2)在不改变样本数量的前提下,提出了一种异常数据自适应修正算法,有效降低了异常数据对聚类性能的影响,且修正后的数据聚类使用加权平

均距离准则比最小距离准则稳定性更强。

(3)为了使基本概率分配更加合理,提出了一种基于加权相似度的函数方法,并将此方法应用在温室环境控制决策融合中。实验对比结果表明,本文提出的方法简单有效,不但能够将证据的不确定程度降低1~2个数量级,而且有效地降低了数据之间的冲突程度,有利于快速做出正确决策。

参 考 文 献

1 李昌玺,周焰,林茵,等. 考虑传感器置信度的改进 D-S 证据合成算法[J]. 解放军理工大学:自然科学版,2017,18(1): 81-86.
LI Changxi, ZHOU Yan, LIN Han, et al. Improved D-S evidence combination rule based on reliability of sensors[J]. Journal of PLA University of Science and Technology: Natural Science Edition, 2017, 18(1): 81-86. (in Chinese)

2 邓雪峰,孙瑞志,聂娟,等. 基于时间自动机的温室环境监控物联网系统建模[J/OL]. 农业机械学报,2016,47(7): 301-308. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20160741&flag=1. DOI:10.6041/j.issn.1000-1298.2016.07.041.
DENG Xuefeng, SUN Ruizhi, NIE Juan, et al. Greenhouse environment monitoring IOT system modeling based on timed automata [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(7): 301-308. (in Chinese)

3 郭文川,程寒杰,李瑞明,等. 基于无线传感器网络的温室环境信息监测系统[J]. 农业机械学报,2010,41(7): 181-185.
GUO Wenchuan, CHENG Hanjie, LI Ruiming, et al. Greenhouse monitoring system based on wireless sensor networks[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(7): 181-185. (in Chinese)

4 何东健,邹志勇,周曼. 果园环境参数远程检测 WSN 网关节点设计[J]. 农业机械学报,2010,41(6): 182-186.
HE Dongjian, ZOU Zhiyong, ZHOU Man. Design of WSN gateway nodes for remote detection of orchards environment parameters [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(6): 182-186. (in Chinese)

5 李莉,李海霞,刘卉. 基于无线传感器网络的温室环境监测系统[J]. 农业机械学报,2009,40(增刊): 228-231.
LI Li, LI Haixia, LIU Hui. Greenhouse environment monitoring system based on wireless sensor network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(Supp.): 228-231. (in Chinese)

6 蔡振江,康健一,张青,等. 数据融合技术在温室温度检测中的应用[J]. 农业机械学报,2006,37(10): 101-103.
CAI Zhenjiang, KANG Jianyi, ZHANG Qing, et al. Application of multi-sensor data fusion in greenhouse temperature test system [J]. Transactions of the Chinese Society for Agricultural Machinery, 2006, 37(10): 101-103. (in Chinese)

7 温静,程茂. 多传感器数据融合技术在温室控制中的应用[J]. 河北农业大学学报,2015,38(4): 111-115.
WEN Jing, CHENG Mao. Application of multi-sensor data fusion technique in greenhouse control[J]. Journal of Agricultural

University of Hebei, 2015, 38(4): 111 – 115. (in Chinese)

8 路盈盈, 陈川, 毛云芳. 基于传感器网络的 K-均值聚类算法研究[J]. 计算机工程与设计, 2007, 28(6): 1349 – 1351.
LUO Yingying, CHEN Chuan, MAO Yunfang. Research on K-means clustering arithmetic based on sensor networks[J]. Computer Engineering and Design, 2007, 28(6): 1349 – 1351. (in Chinese)

9 王俊, 刘刚. 融合粗糙集和证据理论的温室环境控制推理决策方法[J]. 农业工程学报, 2012, 28(17): 172 – 178.
WANG Jun, LIU Gang. Decision-making method based on rough set and evidential theory for greenhouse environmental control [J]. Transactions of the CSAE, 2012, 28(17): 172 – 178. (in Chinese)

10 熊迎军, 沈明霞, 陆明洲, 等. 温室无线传感器网络系统实时数据融合算法[J]. 农业工程学报, 2012, 28(23): 160 – 166.
XIONG Yingjun, SHEN Mingxia, LU Mingzhou, et al. Algorithm of real time data fusion for greenhouse WSN system [J]. Transactions of the CSAE, 2012, 28(23): 160 – 166. (in Chinese)

11 宋庆恒, 刘英德, 马源, 等. 基于多传感器数据融合的温室温湿度控制系统设计[J]. 江苏农业科学, 2015,43(6):394 – 396.

12 李富娟, 李仕华, 赵新远, 等. 温室控制系统多传感器数据融合方法的设计[J]. 湖北农业科学, 2016(16): 4287 – 4289.
LI Fujuan, LI Shihua,ZHAO Xinyuan, et al. Design of multi-sensor data fusion algorithm of greenhouse control system [J]. Hubei Agricultural Sciences, 2016(16): 4287 – 4289. (in Chinese)

13 YANG Y, HAN D, DEZERT J, et al. Comparative study of focal distance measures in theory of belief functions [C] //2016 19th International Conference on Information Fusion, IEEE, 2016:1308 – 1315.

14 SHANNON C E. A mathematical theory of communication. AT&T tech J [J]. ACM Sigmoblie Mobile Computing & Communications Review, 2001, 5(1): 3 – 55.

15 DENG Y. Entropy: a generalized shannon entropy to measure uncertainty[J]. 2015. Available online: <http://vixra.org/abs/1502.0222>.

16 WANG J, XIAO F, DENG X, et al. Weighted evidence combination based on distance of evidence and entropy function[J]. International Journal of Distributed Sensor Networks, 2016, 12(7): 3218784 – 3218784.

17 WANG J, QIAO K, ZHANG Z, et al. A new conflict management method in Dempster – Shafer theory[J]. International Journal of Distributed Sensor Networks, 2017: 1769650.

18 曹建福, 曹雯. 基于改进证据理论的大型制造装备故障诊断[J]. 振动、测试与诊断, 2012, 32(4): 532 – 537.

19 刘海波, 殷越, 艾永乐. 基于 D – S 证据理论的瓦斯突出危险等级评判策略[J]. 济南大学学报: 自然科学版, 2017, 31(1): 73 – 76.
LIU Haibo, YIN Yue, AI Yongle. Evaluation strategy of gas outburst risk level based on D – S evidence theory[J]. Journal of University of Ji'nan: Science and Technology, 2017, 31(1): 73 – 76. (in Chinese)

20 李文立, 郭凯红. D – S 证据理论合成规则及冲突问题[J]. 系统工程理论与实践, 2010, 30(8): 1422 – 1432.
LI Wenli, GUO Kaihong. Combination rules of D – S evidence theory and conflict problem [J]. Systems Engineering-Theory & Practice, 2010, 30(8): 1422 – 1432. (in Chinese)

21 蒋黎明, 何加浪, 张宏. D – S 证据理论中一种新的冲突证据融合方法[J]. 计算机科学, 2011, 38(4): 236 – 238.
JIANG Liming, HE Jialang, ZHANG Hong. New fusion approach for conflicting evidence in D – S theory of evidence [J]. Computer Science, 2011, 38(4): 236 – 238. (in Chinese)

22 安春莲, 黄静, 吴耀云. 基于证据理论的多源信息融合模型[J]. 电子信息对抗技术, 2017, 32(1): 23 – 26.
AN Chunlian, HUANG Jing, WU Yaoyun. A multi-source information fusion model based on evidence theory [J]. Electronic Information Warfare Technology, 2017, 32(1): 23 – 26. (in Chinese)

23 LIU Z, PAN Q, DEZERT J. Classifier fusion based on cautious discounting of beliefs [C] // International Conference on Information Fusion, IEEE, 2016.