

改进证据分类合成方法在农作物生长环境评估中的应用

叶继华 聂小斯

(江西师范大学计算机信息工程学院, 南昌, 330022)

摘要: 针对传统 D-S 证据理论存在处理冲突证据的不足, 基于证据间的相似度引入了信息熵属性, 修正了证据分类属性, 结合证据间相似度属性将证据集重新划分为可信度高证据、一般性证据和冲突证据, 对分类的证据集赋予不同的重要性系数, 并加以修正改进。改进后使得一般性证据和高冲突证据向可信度高的证据意见靠拢, 最后利用 D-S 组合规则对于修正后的证据进行合成。针对农作物生长环境中多个传感器获取的数据构造其所对应证据的基本概率分配函数, 利用模糊理论对基本概率分配函数进行取值。实验采用各类传感器测得的真实数据集进行实验, 结果表明改进的方法既能够很好地解决冲突问题, 同时能降低证据的不确定性。

关键词: 证据分类; 信息熵; 信息融合

中图分类号: TP301 **文献标志码:** A

Improved Evidence Classification Synthetic Method in Environment Assessment of Crop Growth

Ye Jihua, Nie Xiaosi

(Computer Information Engineering College, Jiangxi Normal University, Nanchang, 330022, China)

Abstract: Considering traditional D-S evidence theory deficiencies existing in dealing with conflict evidence, the information entropy attribute is put forward based on the similarity between each evidences, which fixed the evidence classification properties. Combining the similarity attribute between each evidences, the evidence set could be divided into high credibility evidence, general evidence and conflict evidence. The sorted evidence set is given different importance coefficients, and is modified to improve. After modifying, the general evidence and high conflict evidence are closed to the high credibility of evidence opinions. Finally, D-S combination rule is utilized to synthesis for the modified evidence. It is difficult for obtaining data by multiple sensors to establish the basic probability distribution function for the evidence. For the problem, making full use of the ability that rough set theory can deal with incomplete information and knowledge, the decision information table is obtained via the attribute reduction of rough set. The function values are assigned for the basic probability by the decision information table. Combining the rough set and the improved D-S evidence theory, the real data sets are measured by all kinds of sensors. The experimental results show that the improved method can not only effectively solve the conflict problem, but also reduce the uncertainty of evidence.

Key words: evidence classification; information entropy; information fusion

引言

D-S 证据理论^[1]具有表达“不确定”的能力,能够在缺少先验信息的条件下处理不确定性问题。但其缺陷却也非常明显,在其证据冲突时,利用 D-S 组合规则会导致证据失效问题。为了克服 D-S 证据理论的不足,国内外一些知名学者陆续提出不少修正方法,但是文献[2]组合规则在处理低冲突证据的情况下,该组合规则比较可行。但随着证据的增加,证据都逐步地集中于不确定 Θ 集合,使得 $m(\Theta) \rightarrow 1$,这也意味着冲突证据不能提供过多的有效信息,从而该规则不具备一定的实用性。文献[3]在文献[4]的基础上利用相似度矩阵求解除证据所对应的权重,并通过权重求解加权平均取值,虽然经修正后,证据的抗干扰能力更强了,收敛速度更快了,但并不满足证据间的交换律以及结合律。文献[4]改进证据源的组合方法能够使信任度函数的收敛能力增强,但其仅仅是以相同的权重来进行计算的,其得出的结果也并不真实。文献[5]充分考虑了证据间的相识度,使结果更加可信,但未对证据自身信任度进行分析。文献[6]算法规则主要还是计算证据间两两冲突量的平均值来定义有效性系数,将总冲突按比例分为各个命题,但此方法是假定证据的可信度相同,这不合理。文献[7]主要利用证据间的相似属性对证据集进行分类修正,并对于分类后的证据分别赋予不同的修正权值,降低证据间的冲突量,但其将所有不冲突证据的修正系数均赋为 1,从而也导致结果并不精确。改进方法主要分为两类:(1)主要针对证据间的冲突重新分配。(2)主要针对证据的数据源进行修改。本文采用第二种改进方法,基于证据分类的合成算法,提出一种新的证据分类方法进行改进,之后再对各类分类证据集进行相应的权值调整,最后进行 D-S 融合。针对多个传感器获取的数据进行构造其所对应的基本概率分配函数是一个难点。本文利用模糊集理论能够处理不完整信息与知识的能力,结合粗糙集的数据离散化算法得到决策信息表,根据模糊隶属度函数对基本概率分配函数进行取值。结合农作物生长环境的实际应用情况,依据各类型传感器所采集到的真实数据,对农作物生长环境进行评估。

1 D-S 理论和证据属性分析

定义 1: 辨识框架 Θ , 其辨识框架是关于其命题的相互独立可能命题的有限集合^[1]。

定义 2: 基本概率分配函数 $m: 2^\Theta \in [0, 1]$, 其中 2^Θ 为 Θ 的幂集。 $m(\varphi) = 0$, $\sum_{A \subseteq \Theta} m(A) = 1$ 。 m 为辨识框架 Θ 上的基本概率赋值函数^[4], $m(A)$ 为对目标 A 的可信度, 表示对 A 的精确信任度。

定义 3: D-S 规则组合式^[1]: 辨识框架 Θ 下两证据基本概率赋值 (Basic probability assignment, BPA), m_1, m_2 , 则 D-S 组合规则为

$$m(C) = \begin{cases} \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - k} & C \neq \varphi \\ 0 & C = \varphi \end{cases} \quad (1)$$

式中: $C \in 2^\Theta$, $A_i \in 2^\Theta$, $B_j \in 2^\Theta$; 其中 k 表示证据之间的冲突概率

$$k = \sum_{A_i \cap B_j = \varphi} m_1(A_i) m_2(B_j) \quad (2)$$

1.1 证据间相似属性

在多传感器数据融合中,若其中某条证据被其他证据所支持程度大,可得该条证据可信度高,对融合结果的影响较大,反之,则影响较小。证据间相似属性分析利用证据间的距离、冲突、夹角这几种属性进行。设识别框架 Θ 包含 N 个两两不同的命题的完备识别框架, E_1 和 E_2 为 Θ 下的两个证据,其对应的基

本概率分配函数为 m_1 和 m_2 , 焦点分别是 A_i 和 B_j , 两证据之间的距离可以表示为

$$d_{12} = d(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T D(m_1 - m_2)} \quad (3)$$

式中: D 为一个 $2^N \times 2^N$ 的矩阵, 矩阵中的元素为

$$D(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

两证据间的夹角表示为

$$\eta_{12} = \cos\theta = \frac{m_1 m_2}{|m_1| |m_2|} \quad (5)$$

采用 3 个参数作为证据间相似度属性分类的一部分, 主要在于任何一个参数都无法完整地衡量证据的相似度。本文基于证据分类算法, 在证据间相似度属性基础上, 引入证据自身信息熵属性, 重新定义冲突证据、一般证据和可信度高的证据, 并按照新的属性划分证据, 并对冲突证据及一般性证据进行相应的修正, 从而避免了证据划分的不准确性。

1.2 证据信息熵属性

已知 M 个信息源产生的证据, 对于某一假设集 $U = \{\{U_1\}, \{U_2\}, \dots, \{U_N\}\}$, 有 M 条证据 $m_i = (m_{1,i}, \dots, m_{N,i})$, $i = 1, 2, \dots, m$, 结合 D-S 证据理论, 满足 $\sum_{j=1}^N m_{j,i} = 1$ 。

第 i 个证据的信息熵为

$$I_i(m) = - \sum_{j=1}^N m_{j,i} \log m_{j,i} \quad i = 1, 2, \dots, M \quad (6)$$

$I_i(m)$ 为证据 m_i 的信息熵, 信息熵反映了该证据中所包含信息量的大小。当某条证据的信息熵愈大, 可信度越小, 不确定也就愈大, 反之信息熵越小, 所获得不确定性也就越小。

2 基于信息熵的改进证据分类合成方法

在实际应用中, 由于噪声、传感器的不稳定等干扰因素影响, 证据自身也就存在一定的不确定。在数据融合过程中, 每条证据的可信度权重并不完全一样, 而仅仅基于证据间相似度并不能完全表征证据的综合可信度。为此利用信息熵的属性, 对证据的自身可信度进行分析, 将主观因素和客观因素相结合, 综合考虑证据信息熵属性和证据间相似度, 使一般性证据和冲突证据向可信度高的证据意见靠拢, 从而进行更有效、更精确的融合。

2.1 证据间相似性属性计算

证据 E_i 与证据集 E 中其他证据的平均距离为

$$\bar{d}_i = \sum_{j=1, j \neq i}^n d_{i,j} / n - 1 \quad (7)$$

则证据集 E 间的平均距离为

$$\bar{d} = \sum_{i=1}^n \bar{d}_i / n \quad (8)$$

同理求出证据 E_i 与证据集 E 中其他证据的平均冲突

$$\bar{k}_i = \sum_{j=1, j \neq i}^n k_{i,j} / n - 1 \quad (9)$$

证据集 E 间的平均冲突为

$$\bar{k} = \sum_{i=1}^n \bar{k}_i / n \quad (10)$$

求出证据 E_i 与证据集 E 中其他证据的平均方向夹角为

$$\bar{\eta}_i = \sum_{j=1, j \neq i}^n \eta_{i,j} / n - 1 \quad (11)$$

证据集 E 间的平均冲突为

$$\bar{\eta} = \sum_{i=1}^n \bar{\eta}_i / n \quad (12)$$

2.2 证据信息熵属性计算

根据信息熵属性的分析,其第 i 个证据的信息熵为

$$I_i(m) = - \sum_{j=1}^M m_{i,j} \log m_{i,j} \quad (13)$$

式中: j 为目标; M 为目标个数。证据集 E 的平均信息熵为

$$\bar{I}(m) = - \sum_i^n I_i(m) / n \quad (14)$$

对于信息熵较大的证据,可以根据与平均证据比较的偏差,结合该证据的所有目标偏差,计算得出该条证据的标准偏差,再对信息熵较大的证据进行修正。

2.3 证据识别分类

本文利用证据自身的信息熵属性与证据间的相似度属性(距离、冲突量和夹角)将所有证据分为 3 类,分别是冲突证据、可信度高证据和一般性证据。冲突证据仅定义为在证据集中与其他证据相似度低(距离、冲突量均较大和夹角较小)且自身信息熵较小的证据。可信度高证据仅定义为除冲突证据外与其他证据相似度高(距离、冲突向量均较小和夹角较大)且自身信息熵相对较小的证据。一般性证据仅定义为信息熵较大的证据。

2.4 改进分类证据步骤

(1) 选取冲突证据。由于冲突证据其证据自身信息熵可能很小,与其他证据意见严重相悖,所以对于冲突证据的分类,基于证据相似度属性特征和信息熵属性进行分类,很好地避免了证据集中分类所产生的不足。采用计算所有证据 E_i 的 $\bar{d}_i, \bar{k}_i, \bar{\eta}_i, \bar{I}_i$ 以及证据集的 $\bar{d}, \bar{k}, \bar{\eta}, \bar{I}$, 将符合条件 $\bar{d}_i > \bar{d}, \bar{k}_i > \bar{k}, \bar{\eta}_i < \bar{\eta}$ 和 $\bar{I}_i < \bar{I}$ 归为冲突证据的集合 E_{bad} 。

(2) 选取可信度高证据。在基于剩余的证据集,结合证据间相似度属性和证据信息熵属性,计算剩余证据 E_i 的 $\bar{d}'_i, \bar{k}'_i, \bar{I}'_i, \bar{\eta}'_i$ 以及证据集的 $\bar{d}', \bar{k}', \bar{I}', \bar{\eta}'$, 将满足 $\bar{d}'_i < \bar{d}', \bar{k}'_i < \bar{k}', \bar{I}'_i < \bar{I}', \bar{\eta}'_i > \bar{\eta}'$ 的证据划入到可信度高证据集合 E_{good} 中。

(3) 选取一般性证据。相似度属性证据分类^[6]没有结合证据信息熵属性考虑,为了融合的有效性,定义一般性证据是指在自身的信息熵相对较大的证据。满足 $\bar{I}'_i > \bar{I}'$, 符合条件,归为一般性证据集 E_{ord} 。

2.5 基于信息熵的原始证据修正

(1) 可信度高证据的修正方法。为了避免一票否决,若其可信度高证据中的焦元 E_i 满足 $\Theta = \cup E_i$, 则不需要做任何改进,直接融合。否则把不满足可信度高证据条件的证据归类为一般性证据和冲突证据。

(2) 一般性证据修正计算方法。由于证据间的冲突量是验证证据间的基本属性,所以结合各证据对所得信息熵进行权值调整。利用式(5)对证据信息熵进行计算,求一般证据和可信度高证据的信息熵之比以及冲突量之比: $\bar{I} = I_{\text{good}}/I_{\text{ord}}, \bar{k} = k_{\text{good}}/k_{\text{ord}}$, 根据 $\omega_{\text{ord}} = \bar{I}\bar{k}, \bar{k}_{\text{ord}}$ 为一般性证据与可信度高证据的平均冲突量, \bar{k}_{good} 为可信度高证据间的平均冲突量, \bar{I}_{ord} 为一般性证据的平均信息熵, \bar{I}_{good} 同理,再计算权值的证据加权平均值

$$\bar{m}_{\text{ord}\&\text{good}} = \sum_{i=1}^s \omega_i m_i = \left(\sum_{i=1}^s \omega_i m_{1,i}, \dots, \sum_{i=1}^s \omega_i m_{N,i} \right) \quad (15)$$

式中: $\bar{m}_{\text{ord}\&\text{good}}$ 为一般性证据与可信度高证据间的加权证据。 s 为一般性证据与可信度高证据的集合。加权平均证据与第 i 条证据的偏差^[7]表示为

$$\epsilon_i = m_i - \bar{m}_{\text{ord}\&\text{good}} \quad (16)$$

则其偏差可以表示为

$$\epsilon_i = (\epsilon_{1,i}, \dots, \epsilon_{s,i}) \quad (17)$$

m_i 与 $\bar{m}_{\text{ord}\&\text{good}}$ 之间的标准偏差为

$$\delta_i = \sqrt{\sum_{k=1}^N (\epsilon_{k,i})^2 / N} \quad (18)$$

调整后证据为

$$m'_i = (m_{j,i} - \epsilon_{j,i} \delta_i) \quad (19)$$

式中: $j = 1, 2, \dots, N$ 。

(3) 冲突证据修正计算方法。求冲突证据中的 E_i 在改进后的非冲突证据集 $E_{\text{ord}\&\text{good}}$ 上的证据距离 $\bar{d}_{\text{bad},i}$, 冲突 $\bar{k}_{\text{bad},i}$ 和夹角 $\bar{\eta}_{\text{bad},i}$; 再对非冲突证据集里的各证据求其最小的属性参数, $\bar{d}'_{\min} = \min \bar{d}'_i, \bar{k}'_{\min} = \min \bar{k}'_i, \bar{\eta}'_{\max} = \max \bar{\eta}'_i$, 再计算 $d' = \bar{d}'_{\min} / \bar{d}_{\text{bad},i}, k' = \bar{k}'_{\min} / \bar{k}_{\text{bad},i}, \eta' = \bar{\eta}'_{\max} / \bar{\eta}_{\text{bad},i}$ 修正系数为 $\omega'_{\text{bad}} = k' d' \eta'$, 其中 \bar{d}'_{\min} 为非冲突证据集里 \bar{d}'_i 最小的距离, \bar{k}'_{\min} 为非冲突证据集里 \bar{k}'_i 最小的冲突, $\bar{\eta}'_{\max}$ 为非冲突证据集里 $\bar{\eta}_i$ 最大的夹角。

根据权值可信度分配函数调整公式,对冲突证据进行修正

$$\begin{cases} m'_i(A) = \omega_i m_i(A) \\ m'_i(\Theta) = 1 - \omega_i \end{cases} \quad (20)$$

经过式(20)调整,权值小的证据更多地倾向于权值大的证据。综合以上算法,利用式(14~18)对一般性证据进行修正,再对冲突证据赋予其修正系数,最后结合调整后的证据与可信度高性证据进行 D-S 算法规则融合。

2.6 实验测试与分析

实验通过 Matlab 仿真软件进行实验,并利用本文算法与文献[2-4]等的典型算例证据进行比较。已知识别框架 $\Theta = \{a, b, c\}$, 5 组证据 E_1, E_2, E_3, E_4, E_5 均为单命题焦点,假设其基本可信度分配函数如下^[8-9]

$$\begin{aligned} m_1(\{a\}) &= 0.5, m_1(\{b\}) = 0.2, m_1(\{c\}) = 0.3 \\ m_2(\{a\}) &= 0, m_2(\{b\}) = 0.9, m_2(\{c\}) = 0.1 \\ m_3(\{a\}) &= 0.55, m_3(\{b\}) = 0.1, m_3(\{c\}) = 0.35 \\ m_4(\{a\}) &= 0.55, m_4(\{b\}) = 0.1, m_4(\{c\}) = 0.35 \\ m_5(\{a\}) &= 0.55, m_5(\{b\}) = 0.1, m_5(\{c\}) = 0.35 \end{aligned} \quad (21)$$

利用式(7~14)计算出各个证据的 $\bar{d}_i, \bar{k}_i, \bar{I}_i, \bar{\eta}_i$, 以及所有证据集 E 的 $E(\bar{d}, \bar{k}, \bar{I})$ 所得结果如表 1 所示。

表 1 证据集中的参数计算

Table 1 Evidence set parameter calculation

参数	E_1	E_2	E_3	E_4	E_5	$E(\bar{d}, \bar{k}, \bar{I}, \bar{\eta})$
\bar{d}_i	0.221	0.688	0.199	0.199	0.199	0.301
\bar{k}_i	0.648	0.854	0.651	0.651	0.651	0.691
I_i	0.448	0.141	0.403	0.403	0.403	0.360
$\bar{\eta}$	0.832	0.251	0.798	0.798	0.798	0.695

由 2.3 节证据分类可知,满足 $\bar{d}_i > \bar{d}, \bar{k}_i > \bar{k}, \bar{\eta}_i < \bar{\eta}$ 且 $\bar{I}_i < \bar{I}$ 条件的归为冲突证据的集合,由表 1 分析, E_2 证据 $\bar{d}_2 > \bar{d}, \bar{k}_2 > \bar{k}, \bar{\eta}_2 < \bar{\eta}, \bar{I}_2 < \bar{I}$, 由分类算法,可把 E_2 证据划归为得到冲突证据 $E_{\text{bad}} = \{E_2\}$ 。

基于剩余的证据集,计算剩余非冲突证据 E_i 的 $\bar{d}'_i, \bar{k}'_i, \bar{I}'_i, \bar{\eta}'_i$ 以及证据集的 $\bar{d}', \bar{k}', \bar{I}', \bar{\eta}'$, 由分类 $\bar{d}'_i < \bar{d}', \bar{k}'_i < \bar{k}', \bar{I}'_i < \bar{I}'$ 和 $\bar{\eta}'_i > \bar{\eta}'$, 由表 2 分析可知,证据 $\{E_3, E_4, E_5\}$ 可划为可信度高证据 $E_{\text{good}} = \{E_3, E_4, E_5\}$, 并由信息熵属性 $\bar{I}'_i > \bar{I}'$, 可得到 $\{E_1\}$ 为一般性证据 $E_{\text{ord}} = \{E_1\}$ 。

修正算法对一般性证据和冲突证据进行修正,使得一般性证据向可信度高证据靠拢,并对冲突证据进行修正,从而达到更好的融合效果,对比如表 3 所示。由表 3 可知,D-S 规则对于冲突的证据并不能达到有效的融合,整个融合过程 $m(a) = 0$,直接把 a 否定;文献[2]方法得到证据 $m(a) = 0$ 且 $m(\Theta) \rightarrow 1$;

表 2 非冲突证据集的参数计算

Table 2 Parameter Calculation of non-conflicting evidence collection

参数	E_1	E_3	E_4	E_5	$E(\bar{d}', \bar{k}', \bar{I}', \bar{\eta}')$
\bar{d}'_i	0.087	0.029	0.029	0.029	0.042 5
\bar{k}'_i	0.6	0.577	0.577	0.577	0.583
\bar{I}'_i	0.448	0.403	0.403	0.403	0.414
$\bar{\eta}'_i$	0.984	0.995	0.995	0.995	0.992

表 3 几种典型改进方法比较

Table 3 Comparison of several typical improvement methods

规则	m_1, m_2, m_3	m_1, m_2, m_3, m_4	m_1, m_2, m_3, m_4, m_5
文献[1]	$m(a) = 0, m(b) = 0.631 6$	$m(a) = 0, m(b) = 0.032 88$	$m(a) = 0, m(b) = 0.122 8$
规则	$m(c) = 0.368 4$	$m(c) = 0.671 2$	$m(c) = 0.877 2$
文献[2]	$m(a) = 0, m(b) = 0.018$	$m(a) = 0, m(b) = 0.001 8$	$m(a) = 0, m(b) = 0.02$
规则	$m(c) = 0.010 5, m(\Theta) = 0.971 5$	$m(c) = 0.003 7, m(\Theta) = 0.994 5$	$m(c) = 0.001 3, m(\Theta) = 0.998 5$
文献[4]	$m(a) = 0.35, m(b) = 0.522 4$	$m(a) = 0.6027, m(b) = 0.262 7$	$m(a) = 0.795 8, m(b) = 0.093 2$
规则	$m(c) = 0.127 6$	$m(c) = 0.134 6$	$m(c) = 0.111$
文献[6]	$m(a) = 0.16, m(b) = 0.201$	$m(a) = 0.194, m(b) = 0.16$	$m(a) = 0.211, m(b) = 0.138$
规则	$m(c) = 0.125, m(\Theta) = 0.486$	$m(c) = 0.137, m(\Theta) = 0.509$	$m(c) = 0.144, m(\Theta) = 0.507$
文献[3]	$m(a) = 0.486, m(b) = 0.348$	$m(a) = 0.777, m(b) = 0.063$	$m(a) = 0.891, m(b) = 0.009$
规则	$m(c) = 0.166$	$m(c) = 0.16$	$m(c) = 0.1$
文献[9]	$m(a) = 0.501 8, m(b) = 0.148 1$	$m(a) = 0.688 9, m(b) = 0.062 2$	$m(a) = 0.794 2, m(b) = 0.018$
规则	$m(c) = 0.265 2, m(\Theta) = 0.084 9$	$m(c) = 0.244 2, m(\Theta) = 0.004 8$	$m(c) = 0.187 0, m(\Theta) = 0.000 8$
文献[10]	$m(a) = 0.686 9, m(b) = 0.050 5$	$m(a) = 0.795 8, m(b) = 0.010 6$	$m(a) = 0.8614, m(b) = 0.002 1$
规则	$m(c) = 0.170 1$	$m(c) = 0.193 6$	$m(c) = 0.133 8$
本文	$m(a) = 0.69, m(b) = 0.061$	$m(a) = 0.803, m(b) = 0.013$	$m(a) = 0.867 9, m(b) = 0.003 3$
规则	$m(c) = 0.249$	$m(c) = 0.184$	$m(c) = 0.128 8$

文献[4]规则能够识别目标,但精度并不最优;文献[6]算法规则虽然组合规则很好,但组合效率慢,需要多个证据结合才能得到不错的效果;文献[3]算法规则效果虽好,但其不满足融合的交流律和结合律;文献[9]算法规则考虑了证据的不确定性,但其通过加权后的证据来计算不确定性,带有一定的主观性;文献[10]算法规则并未对信息的可信度、不确定性加以考虑;本文同样对证据集进行3种类型证据划分,对于一般性证据,其自身信息量的不确定性较高,采用证据间的标准偏差对于其证据进行修正,并向可信度高的证据靠拢,而冲突性证据直接采用折扣系数修正的方法进行改进。综上所述对比实验结果,表明本文方法在识别目标时优于其他算法。

3 改进证据分类合成方法应用

本文以温室大棚中黄瓜为例,利用模糊理论和改进D-S证据理论对黄瓜的生长环境进行评估,以确定黄瓜环境的好坏。本文主要基于温室大棚对黄瓜生长环境评估的4种属性并结合改进的D-S证据理论的评估关系的融合模型。首先由样本数据根据粗糙集理论的数据离散化得到决策性信息表,其次根据模糊统计的隶属度计算出各传感器属性对应的基本概率分配函数。考虑到证据理论融合中的证据冲突问题,结合改进证据分类的合成算法,进行D-S融合,最终得到对于黄瓜生长环境的优良、一般、较差的属性评估。

实验通过各类型传感器采集的数据,结合Rosetta软件实现离散化处理,对于数据离散化,根据属性值由小到大顺序对决策表中实例进行排序,然后判断,对于两个相邻实例,在属性值和决策值都不同的情况下,选取两属性值的平均值作为断点值,从而为所有连续属性找到合适的断点集。其求解基本概率分配函数的具体过程如下:

- (1) 首先需要对传感器测得的原始数据进行处理。假设输入 M 个不同信息源,样本取值 N 组,则该决策表应为 N 行, $M+1$ 列,其中决策表的 M 列代表条件属性,而最后一列代表决策属性。
- (2) 对于输入的样本信息进行离散化处理,并最终形成决策表。
- (3) 对比样本数据决策表,根据决策数据的取值,利用模糊统计法理论构建得到隶属度函数与基本概率赋值。

基于种植经验,以黄瓜结果期温度为实例,粗糙集离散化算法处理,分别得到4个属性证据的断点以及断点波动温度 a ,同理某时刻适宜的湿度 b 、土壤温度 c ,空气 CO_2 浓度 d ,根据不同取值将连续区间离散化为定性区间:生长优良、生长一般、生长较差3部分。离散化后的决策表如表4所示。

表4 各属性决策一览表
Table 4 Decision list properties

属性证据	生长优良	生长一般	生长较差
空气温度属性	$(a-2, a+1)$	$(a-3, a-2) \cup [a+1, a+3)$	$(-\infty, a-3] \cup [a+3, +\infty)$
空气湿度属性	$(b-20, b+10)$	$(b-30, b-20) \cup [b+10, b+25)$	$(-\infty, b-30] \cup [b+25, +\infty)$
土壤温度属性	$(c-1, c+1)$	$(c-2, c-1) \cup [c+1, c+2)$	$(-\infty, c-2] \cup [c+2, +\infty)$
CO_2 浓度属性	$(d-200, d+200)$	$(d-300, d-200) \cup [d+200, d+400)$	$(-\infty, d-300] \cup [d+400, +\infty)$

对于D-S证据理论融合问题首要任务是如何有效地构造基本概率分配函数。而隶属度是指各种已知性能评价指标隶属于特定评价等级的概率,其能够对一个确定元素是否属于一个可变动的集合做出准确判断。因此,利用隶属度计算基本信任分配可以较好地处理模糊性及主观判断等问题。在隶属度的计算方法中,模糊统计法可以较直观地反映模糊概念中的隶属程度。

采用基于模糊统计法计算待评估的生长情况的基本信任概率分配函数,其可分为如下步骤:

- (1) 计算当前属性证据,待评估的生长情况为优良的数量与总的评估数量之比。得到该比值即为评估生长情况优良的隶属程度,可以表示为: $u(A)$,同理利用模糊统计法,得到生长情况为一般和较差的

隶属程度,分别表示 $u(B), u(C)$ 。

(2)在基于获取的属性证据的基本概率分配函数,该取值等于隶属度为优良与隶属度为一般和较差的三者之和的比值,从而基本概率分配函数可以记作 $m(A)$,同理得到隶属度为一般和较差的基本概率分配函数,分别记作 $m(B)$ 和 $m(C)$ 。

选取合适的传感器,对于空气温度、土壤温度、 CO_2 浓度、空气湿度选取相应的传感器类型进行测试。根据第3节的模糊统计法的求解,得到各传感器对于评估属性的隶属度函数。由式 $m(A) = \frac{u(A)}{u(A) + u(B) + u(C)}$,同理求出 $m(B)$ 和 $m(C)$,可得到基本概率分配函数如下

$$\begin{aligned} m_1(A) &= 0.51, m_1(B) = 0.19, m_1(C) = 0.3 \\ m_2(A) &= 0.6, m_2(B) = 0.1, m_2(C) = 0.3 \\ m_3(A) &= 0.6, m_3(B) = 0.1, m_3(C) = 0.3 \\ m_4(A) &= 0.6, m_4(B) = 0.1, m_4(C) = 0.3 \end{aligned} \quad (22)$$

利用 D-S 证据理论,最终融合

$$m(A) = 0.93, m(B) = 0.0016, m(C) = 0.0684 \quad (23)$$

由以上计算过程可知,在无冲突证据的情况下,该修正方法相比原始 D-S 组合规则有一定的调整。而在产生冲突证据的情况下,基本概率分配函数如下所示

$$\begin{aligned} m_1(A) &= 0.51, m_1(B) = 0.19, m_1(C) = 0.3 \\ m_2(A) &= 0, m_2(B) = 0.2, m_2(C) = 0.8 \\ m_3(A) &= 0.6, m_3(B) = 0.1, m_3(C) = 0.3 \\ m_4(A) &= 0.6, m_4(B) = 0.1, m_4(C) = 0.3 \end{aligned} \quad (24)$$

利用 D-S 证据理论,最终融合结果对比如表 5 所示。由以上步骤,可以由黄瓜生长的 4 种生长属性,根据模糊统计试验方法获取到隶属度函数,再接着获取属性下证据的基本概率分配函数,隶属度为生长优良与隶属度为优良、一般和较差三者之和的比值。由此方法,可获取多传感器以各自属性为标准判定目标的基本概率分配函数,从而为实际应用的评估决策打下坚实基础。

表 5 实际黄瓜生长环境评估的对比

Table 5 Actual cucumber growing environmental assessment comparison

规则	m_1, m_2, m_3	m_1, m_2, m_3, m_4
文献[1]规则	$m(A) = 0, m(B) = 0.143, m(C) = 0.857$	$m(A) = 0, m(B) = 0.053, m(C) = 0.947$
文献[2]规则	$m(A) = 0, m(B) = 0.004,$ $m(C) = 0.072, m(\Theta) = 0.924$	$m(A) = 0, m(B) = 0.004, m(C) = 0.0216$ $m(\Theta) = 0.978$
文献[6]规则	$m(A) = 0.171, m(B) = 0.082,$ $m(C) = 0.29, m(\Theta) = 0.457$	$m(A) = 0.216, m(B) = 0.077$ $m(C) = 0.238, m(\Theta) = 0.469$
文献[10]规则	$m(A) = 0.7228, m(B) = 0.0518$ $m(C) = 0.2254$	$m(A) = 0.8563, m(B) = 0.0102$ $m(C) = 0.1335$
本文规则	$m(A) = 0.7328, m(B) = 0.0482$ $m(C) = 0.219$	$m(A) = 0.8618, m(B) = 0.0094$ $m(C) = 0.1288$

4 结束语

本文利用基于温室大棚对黄瓜生长环境评估的 4 种属性并结合改进的 D-S 证据理论的评估关系融合模型,最后结合 D-S 组合规则融合。改进的算法更好地解决了证据间的高冲突悖论问题以及不确定问题,在温室大棚黄瓜的生长评估中起到更好的决策作用,并对于农作物产量的提高具有极大的意义。

参考文献:

- [1] Dempster A P. Upper and lower probabilities induced by a multi-valued mapping[J]. *Annual Math Statist*, 1967, 38(4):325-339.
- [2] Yager R R. On the Dempster-Shafer framework and new combination rules[J]. *Information Science*, 1989, 41(2):93-137.
- [3] 邓勇, 施文康. 一种有效处理冲突证据的组合方法[J]. *红外与毫米波学报*, 2004, 23(1):27-32.
Deng Yong, Shi Wenkang. Efficient combination approach of conflict evidence[J]. *J Infrared Millim*, 2004, 23(1):27-32.
- [4] Murphy C K. Combining belief function when evidence conflicts[J]. *Decision Support Systems*, 2000(29):1-9.
- [5] 刘希亮, 陈桂明, 李方溪. 基于距离测度的证据合成方法[J]. *数据采集与处理*, 2014, 29(1):121-125.
Liu Xiliang, Chen Guiming, Li Fangxi. Approach to evidence combination based on distance measurement[J]. *Journal of Data Acquisition and Processing*, 2014, 29(1):121-125.
- [6] 孙全, 叶秀清. 一种新的证据组合规则[J]. *电子学报*, 2000, 28(8):117-119.
Sun Quan, Ye Xiuqing. A new combination rules of evidence theory[J]. *Acta Electronica Sinica*, 2000, 28(8):117-119.
- [7] 彭颖, 沈怀荣, 马永一. 一种新的冲突证据融合方法[J]. *兵工学报*, 2011, 32(1):79-85.
Peng Ying, Shen Huairong, Ma Yongyi. A new fusion method for conflicting evidence[J]. *Acta Armamentarii*, 2011, 32(1):79-85.
- [8] 战红, 谭继文, 薛金亮. 基于信息熵与判断矩阵的 D-S 证据理论改进方法在故障诊断中的应用[J]. *北京工业大学学报*, 2013, 39(8):1140-1143.
Zhan Hong, Tan Jiwen, Xue Jinliang. Application of an improved D-S evidence theory based on information entropy and evaluation matrix to fault diagnosis[J]. *Journal of Beijing University of Technology*, 2013, 39(8):1140-1143.
- [9] 熊彦铭, 杨占平. 基于模型修正的冲突证据组合方法[J]. *控制与决策*, 2011, 26(6):883-887.
Xiong Yanming, Yang Zhanping. Novel combination method of conflict evidence based on evidential model modification[J]. *Control and Decision*, 2011, 26(6):883-887.
- [10] 王亮, 吕卫民, 滕克难, 等. 基于分类修正的多证据合成方法[J]. *控制与决策*, 2015, 30(1):125-130.
Wang Liang, Lü Weimin, Teng Kenan, et al. Combination method of multi-evidence based on classification correction[J]. *Control and Decision*, 2015, 30(1):125-130.

作者简介:



叶继华(1966-),男,教授,
研究方向:物联网技术, E-
mail: yjhwcl@163.com。



聂小斯(1991-),男,硕士研
究生,研究方向:物联网技
术。