

DOI: 10.7652/xjtuxb201804021

一种可视化的加权平均信息融合方法

卫博雅, 寿业航, 蒋雯
(西北工业大学电子信息学院, 710072, 西安)

摘要: 针对传统加权信息融合方法不能有效降低信息损失、快速做出合理判断的问题,提出了一种可视化的加权平均信息融合方法。首先,将原始信息数据按数值大小进行降序排列,生成一个有序序列,再采用可视图转换法将该有序序列的数据转化到网络拓扑结构中,其中数据和网络节点一一对应;然后,利用支持度函数对网络中的节点关系进行衡量,从而得到原始数据的权重;最后,将原始数据进行加权融合,得到最终结果。与经典的融合方法相比,新方法从一个更客观的角度对数据权重进行处理,准确度更高,对数据融合及决策问题的处理更有效。实验结果表明:新方法具有更好地实际可操作性和合理性;更重要的是,新方法降低了融合过程中的信息损失,几乎不会产生与直觉相悖的融合结果,大大提高了人们在决策分析处理中的有效性。

关键词: 信息融合;可视图;复杂网络;融合算子;决策分析

中图分类号: TP391.9 **文献标志码:** A **文章编号:** 0253-987X(2018)04-0145-05

A Visibility Graph Information Fusion Method with Weighted Average

WEI Boya, SHOU Yehang, JIANG Wen
(School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: A visibility graph fusion method with weighted average is proposed to solve the problem that the classical weighted averaging fusion methods cannot effectively reduce the information loss and make reasonable judgments quickly. First, argument values are ordered in a descending order to form a time series. After that, a network is constructed based on the visibility graph, and the time series is converted into a graph. The data and network nodes correspond one to one. Then a support function is proposed to measure the relationship between the nodes, and the weights of original data are obtained. Finally, the original data are weighted and fused to get final results. Compared with the classical methods, the new method deals with the weights from a more objective perspective, and it is more accurate and effective when dealing with decision making problems. An example of produced water management is presented to show that the proposed method is reasonable and practical. Moreover, this method reduces the information loss in the fusion process, and hardly produces a counter-intuitive result. It greatly improves the effectiveness of decision analysis.

Keywords: information fusion; visibility graph; complex networks; aggregation operator; decision analysis

信息融合最早称为数据融合,是一个对从单个和多个信息源获取的数据和信息进行关联、相关和综合,以获得精确的位置和身份估计,以及对态势和威胁及其重要程度进行全面及时评估的信息处理过程。该技术起源于20世纪70年代,并在90年代迅速发展起来,广泛应用于各个领域。然而,目前该技术仍然存在精确度不高、融合效率低等问题,需要进一步研究处理。信息融合通常采用融合算子对数据进行相关融合。而决策问题的多种多样促使人们提出了大量的融合方法,这些方法试图为不同的决策问题提供相应的解决方案。

融合算子是信息融合领域一种十分有用的工具,它可以提供更多功能的数据融合过程。例如,Yager于1988年首先提出了有序加权平均(ordered weighted averaging operator,OWA)算子^[1]。这种有序加权方法可以有效消除一些不合理的信息融合情况。此后,该算子的衍生算子被大量提出,如被诱导的OWA算子(IOWA)^[2]、不确定OWA算子(UOWA)^[3]、广义OWA算子(GOWA)^[4]等。此后,Yager从数据本身的关系入手,进一步提出了幂均(power average aggregation operator,P-A)算子^[5],该算子包含了一个支持度函数,为了反映融合过程中数据间相互支持的原则,幂均算子引入了一种新的支持度函数。

自OWA算子首次出现以来,专家学者们对其权重的获取进行了大量的研究。O'Hangan在1988年提出了获得权重的关键方法之一——最大熵方法^[6],该方法将OWA算子权重问题定义为约束的非线性优化模型,并以orness的预定度作为目标函数,得到的权重被称为最大熵权重。之后Fullér和Majlender采用拉格朗日乘子法^[7]对O'Hangan^[6]的问题进行了分析。然而,如何确定权重迄今仍是一个有待讨论的问题^[8]。

不仅如此,Yager于2001还首次提出了幂均算子^[5]。不同于OWA算子,在该算子中数据的权重是由其他数据给它的支持度决定的。幂均算子广泛应用于信息融合^[9]、环境评估^[10]和多属性决策^[11-12]等领域。近年来,学者们对P-A算子进行了深入研究,衍生出许多不同的算子来更好地处理不确定信息。例如,直觉模糊幂集合算子(IFPA)^[12]、有序加权几何幂均算子(POWGA)和有序加权调和幂均算子(POWHA)^[13]。然而,这些基于幂均算子的融合方法过于复杂,在实际应用中无法做到简单有效。为了方便地解决实际问题,本文基于幂均算子,充

分考虑了数据间的相互关系提出了一种新的可视化加权平均信息融合方法,该方法通过数据之间的相互支持度来衡量数据的权重关系,并利用该方法获得的权重对原始数据进行加权融合,以针对不同问题作出相应决策。

1 幂均算子

幂均算子首先由著名学者Yager于2001年提出^[5]。该算子包含了一个支持度函数,此函数使数据在融合过程中应该相互支持。

定义1 设 $(a_1, a_2, \dots, a_n)$ 为传感器输出的一组原始信息数据,则幂均算子 F 的定义如下

$$F(a_1, a_2, \dots, a_n) = \frac{\sum_{i=1}^n (1 + T(a_i)) a_i}{\sum_{i=1}^n (1 + T(a_i))} \quad (1)$$

式中

$$T(a_i) = \sum_{\substack{j=1 \\ j \neq i}}^n S(a_i, a_j) \quad (2)$$

$T(a_i)$ 为其他数据对 a_i 的所有支持度之和; $S(a_i, a_j)$ 为对 a_i 的支持度。 $S(a_i, a_j)$ 满足以下3个性质:

- (1) $S(a_i, a_j) \in [0, 1]$;
- (2) $S(a_i, a_j) = S(a_j, a_i)$;
- (3) 如果 $|a_i - a_j| < |a_x - a_y|$,那么 $S(a_i - a_j) \geq S(a_x - a_y)$ 。

根据定义可以看出,幂均算子是一个非线性加权平均算子,数据的排列顺序并不影响最终结果值。如果两个数据越相似,二者的支持度就越大。当 $S(a_i, a_j) = 0$ 时,幂均算子将退化为算数平均算子。

2 可视图理论

可见图(Visibility Graph, VG)方法是由Lacasa等人于2008年首次提出^[14],该方法可以将一个时间序列转换成一幅可视图。在图中,每个节点和序列数值一一对应,图中节点间的连接反映了数据之间的可见性。可视图的定义如下。

定义2 设坐标上的两个数据 (i, a_i) 和 (j, a_j) 之间具有可见性,那么对于任意一个在二者间的数值 (k, a_k) 满足

$$a_k < a_j + (a_i - a_j) \frac{j - k}{j - i} (i, a_i) \quad (3)$$

3 可视化信息融合方法

定义3 设 $(a_1, a_2, \dots, a_n)$ 为传感器输出的一组原始

信息数据,将其按照数值大小进行降序排序,得到 $O = \{o_1, o_2, \dots, o_n\}$, 称其为原始序列对应的有序序列。

定义 4 设两个排好序的坐标值分别是 (i, o_i) 和 (j, o_j) , 如果二者间的任意一个变量 (k, o_k) 满足

$$o_k < o_j + (o_i - o_j) \frac{j - k}{j - i} \tag{4}$$

那么称坐标点 (i, o_i) 和 (j, o_j) 具有可见性。

定义 5 设 $O = \{o_1, o_2, \dots, o_n\}$ 是一个有序序列, 那么 n 元函数 F (融合值) 可定义为

$$F(o_1, o_2, \dots, o_n) = \frac{\sum_{i=1}^n (1 + T(o_i)) o_i}{\sum_{i=1}^n 1(1 + T(o_i))} \tag{5}$$

式中

$$T(o_i) = \sum_{\substack{j=1 \\ j \neq i}}^n S(o_i, o_j) \tag{6}$$

在复杂网络中,度是指在无向图中每个节点连边的条数。在一个网络拓扑结构中,节点的度可以代表该点的重要程度。各个的节点度的散布情况就是度的分布。由于度的分布可以反映节点的重要性,一个节点的度越多,那么它受其他节点的支持程度就越大,因此可以通过节点间的距离来定义节点间的支持度,即 $S(o_i, o_j)$ 。

定义 6 设 d 为 2 个节点之间的距离,则二者的支持度可以表示为

$$S(o_i, o_j) = \frac{1}{d_{ij}^n} \tag{7}$$

式中: n 为任意一个正整数。

定义 7 o_i 的支持度总和可以表示为

$$T(o_i) = \sum_{\substack{j=1 \\ j \neq i}}^n S(o_i, o_j) = \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{d_{ij}^n} \tag{8}$$

在可视图,数据的顺序决定了节点与节点之间是否具有可见性。如果两个节点存在可见性,就将二者连接起来。在网络中,如果一个节点与其它节点的连线越多,该节点受其他节点的支持度就越大,对融合结果的影响也越大。

4 案例分析

4.1 仿真算例

设多源传感器输出一组数据,这 8 个数值分别是 45、85、50、70、55、40、70、75。现在要通过这些信息对传感器做一简单评估。

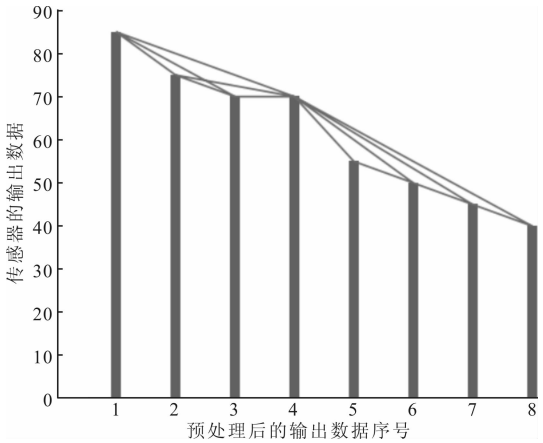
首先,要对其数据进行将序排序,得到一组有序序列含有 8 个数值,即 $O = \{85, 75, 70, 70, 55, 50,$

$45, 40\}$ 。将该序列转化成条形图,并将可视顶点连接起来,构成一副可视图。图 1 十分清晰地展现了条形顶点间的关系,通过该图可以很容易获得每个顶点的支持度总和。例如数值 85,和它有连接关系的 3 个数值分别是 75,75,和 70。选取 $n=2$,根据式(8),可得到第一个点 85 的支持度总和 $T(o_1)$ 为

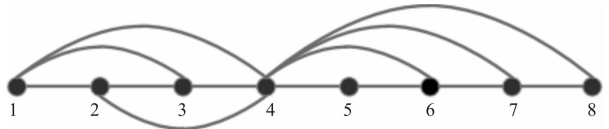
$$T(o_1) = \frac{1}{d_{12}^2} + \frac{1}{d_{13}^2} + \frac{1}{d_{14}^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} = \frac{49}{36} \tag{9}$$

同理,可以得到其他节点的总支持度。之后,根据式(5)可以得到最终的融合结果

$$F(85, 75, 70, 70, 55, 50, 45, 40) = 61.774\ 4 \tag{10}$$



(a)有序序列的可视图



(b)与图 1a 对应的网络结构

图 1 输出数据的可视图及网络结构

4.2 应用案例

可持续发展和环境保护要求具备绿色产品、绿色流程和废物管理等策略。绿色清洁流程与绿色产品的设计及选择会涉及一系列待处理的与环境、经济和技术相关的数据,因此在不确定性环境下,采用综合技术去处理这些因素是十分必要的。

可视幂均融合方法将用于近海岸生产水管理分析。同时,用最大熵 OWA 算子^[8]和 OVGWA 算子^[15]进行计算,并通过融合结果对 3 种方法子的优劣性进行分析。在第一阶段,用最佳可行性技术 (Best Available Technologies, BAT) 对生产水的处理进行评估。本文只考虑第一阶段的评估问题。由

于难以理解这些数据的内在关系,因此处理离散的定量、定性输入数据是十分困难的。这就导致人们往往无法做出正确合理的决策,而本文的可视幂均融合方法则可以有效解决这个难题。

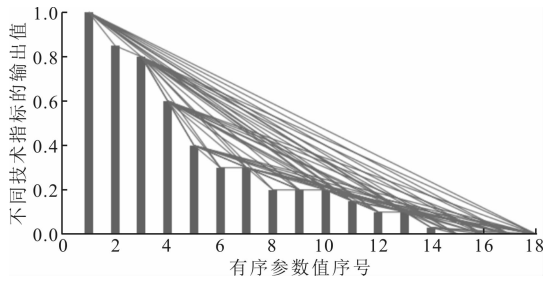
在绿色和清洁工艺的设计和选择中,评估不同的选择。每种选择都会导致不同的结果。在第一阶段的评估中,文献[16]对 14 种不同的 BAT 选择的 18 种技术指标分别进行单独评估,得到离散量化、定性输入数据作为本算例的原始数据(如文献[16]表 4 所示)。这 14 种水处理方式分别是:①物理分离技术包括浮选、喷射、合并;②增强分离技术由水力旋流器、PECT-F 型设备、离心机完成;③抛光技术包括 MPPE 塑胶、吸附、C-Tour 技术、薄膜、蒸汽剥离、生物、生产水再注射、井下分离。

本文方法选取支持度参数 $n=1$,计算此种情况下对决策结果的影响。以浮选技术的各种指标为例,根据文献[16]中的表 4 可得,18 个原始数据值 $\{a_1, a_2, \dots, a_{18}\}$ 为 $\{0.03, 0.10, 0.15, 0.30, 0.20, 0.60, 0.00, 0.00, 0.20, 0.00, 0.00, 0.80, 0.85, 1.00, 0.40, 0.10, 0.20, 0.30\}$,将其排序便可得到原始数据的有序序列,即 $\{o_1, o_2, \dots, o_{18}\}$ 。然后,画出 $n=1$ 时这些数据的相应的可视图,如图 2 所示。根据式(8)可求得每种技术所获取的支持度总和。以此类推,最终根据式(5)即可得到最终融合结果为 0.291 5。同理可得其他 13 种选择的相应融合结果,分别是 0.329 4、0.335 2、0.409 8、0.409 0、0.387 7、0.526 2、0.512 2、0.575 6、0.602 5、0.515 2、0.511 8、0.643 3、0.657 2。

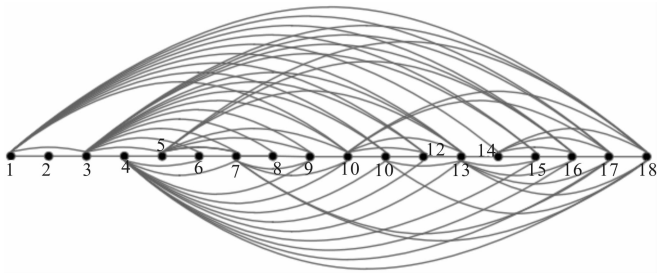
4.3 比较与分析

本文针对不同类别的处理技术(物理和增强分

离、抛光技术)都进行了分析与计算,结果分析与比较如表 1 所示。为了便于观察比较,本文将不同方法得到的融合结果绘制成折线图,如图 3 所示。



(a)对应指标的可视图



(b)与图 2a 对应的网络结构

图 2 物理分离技术的可视图及网络结构

总的来说,不同融合方法得到的融合曲线变化趋势大体一致。对 OWA 算子融合方法来说, α 值会对结果产生重要影响。不同的 α 值会导致不同的融合结果。由于 α 的选取受主观因素的影响非常大,因此选取合适的 α 值是十分困难的。本例中,选取 $\alpha=0.1, 0.5$ 和 0.9 进行融合计算。对于同样用到可视化方法的 OVGWA 算子融合方法来说,结果是符合实际的,但该方法并没有充分利用数据关系

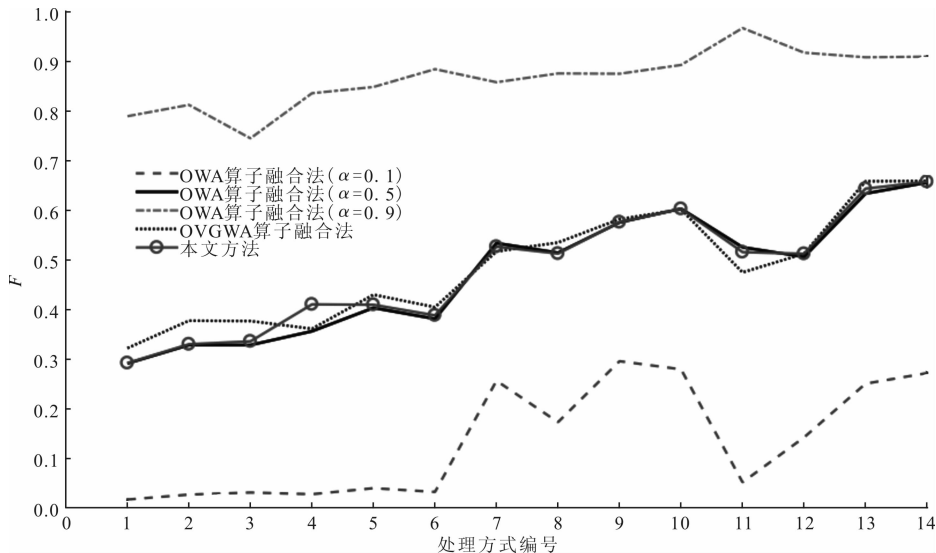


图 3 3 种方法的融合值比较

的紧密度。本文方法的变化曲线和 OVGWA 算子融合方法曲线很接近,这说明了本文方法的合理性。然而,本文认为在网络中衡量节点间的支持度大小时,节点的远近直接影响到它们之间的依赖关系。因此,在设计方法时,考虑节点间距离是十分必要的。OVGWA 算子融合方法认为网络中不同距离的节点支持度是一样的,这显然是不合理的。

根据对不同技术融合值的最大最优选择原则,从表 1 可以看出,本文方法认为融合值为 0.657 2 的井下分离技术是最理想的选择,而在实际的评估

中,这种选择被证实确实是最好的。因此,本文方法具有很强的可操作性和有效性。由于本文方法具有无序性,因此无论原始数据顺序如何改变,结果始终是一样的,这很好地减少了原始数据在融合过程中的信息损失。从图 3 可以看出,本文方法产生的曲线和 OWA 算子融合方法中 $\alpha=0.5$ 的曲线最接近($\alpha=0.5$ 是 OWA 算子融合方法中包含主观因素最少的选择)。造成二者差异的主要原因是数据之间的支持度和权重的不同。通过与不同融合方法进行比较,这个例子证明了本文方法的可行性和有效性。

表 1 3 种方法的融合值比较

编号	处理方式	OWA 算子融合法 ^[8]			OVGW 算子融合法 ^[15]	本文方法
		$\alpha=0.1$	$\alpha=0.5$	$\alpha=0.9$		
1	浮选	0.016 1	0.290 6	0.789 3	0.321 5	0.291 5
2	喷射	0.026 1	0.327 8	0.812 2	0.376 9	0.329 4
3	合并	0.030 3	0.327 7	0.744 7	0.376 2	0.335 2
4	水力旋流器	0.026 9	0.355 6	0.835 7	0.360 6	0.409 8
5	PECT-F 型设备	0.039 0	0.402 8	0.848 3	0.429 4	0.409 0
6	离心机	0.031 5	0.380 6	0.884 3	0.404 2	0.387 7
7	MPPE 塑胶	0.255 7	0.533 3	0.857 9	0.516 5	0.526 2
8	吸附	0.172 2	0.513 9	0.875 6	0.534 6	0.512 2
9	C-Tour 技术	0.295 2	0.575 0	0.875 0	0.581 7	0.575 6
10	薄膜	0.278 8	0.602 8	0.892 7	0.601 4	0.602 5
11	蒸汽剥离	0.050 9	0.525 0	0.967 2	0.474 3	0.515 2
12	生物	0.141 5	0.505 6	0.917 5	0.512 0	0.511 8
13	生产水再注射	0.249 9	0.633 3	0.908 2	0.658 1	0.643 3
14	井下分离	0.271 2	0.655 6	0.909 5	0.658 8	0.657 2

5 结 论

针对传统融合方法存在的种种问题,本文从网络的角度出发,提出了一种有序可视的信息融合方法。首先,将一组数据按照降序顺序排列;然后,将这组数据转换成一个有序可视图,并将顶点具有可见性的条形连接起来,映射到网络中,根据节点间的距离,就可以算出具有可见性的节点之间的相互支持度;最后根据支持度的不同,对所有实验数据进行加权融合,得到最终结果。近海岸生产水管理的实验结果也说明了新方法的合理性和有效性。

本文提出的可视化加权平均信息融合方法有 3 个优点:首先,它充分考虑了一组数列的序信息;其次,从网络拓扑图的角度出发,用数据间的距离来衡量彼此间的关系是十分有趣和新奇的;最后,由于它的可操作性和有效性,未来它将应用于诸如数据挖

掘,智能决策以及信息融合等更多领域。

参考文献:

[1] YAGER R R. On ordered weighted averaging aggregation operators in multi-criteria decision making [J]. Readings in Fuzzy Sets for Intelligent Systems, 1993, 18(1): 80-87.

[2] 王晓,刘兮,陈华友,等. 基于 IOWA 算子的区间组合预测方法 [J]. 武汉理工大学学报(信息与管理工程版), 2010, 32(2): 221-225.

WANG Xiao, LIU Xi, CHEN Huayou, et al. An interval combination forecasting method based on IOWA operators [J]. Journal of Wuhan University of Technology (Information and Management Engineering), 2010, 32(2): 221-225.

(下转第 157 页)