

基于模糊因果聚类预测法预测污水净化程度的研究

蒋莹莹

宿州学院数学与统计学院, 安徽 宿州
Email: right067@126.com

收稿日期: 2020年12月28日; 录用日期: 2021年1月26日; 发布日期: 2021年2月2日

摘 要

研究污水净化程度处理规律, 改进静态模糊预测中因果聚类分析模糊特征的提供方式, 建立新的模糊特征集, 构造对应的正态模糊数预测模型; 通过实例实测数据进行因果聚类分析, 验证模型。

关键词

静态模糊预测, 因果聚类分析, 模糊特征集, 正态模糊数

Research on Prediction of Sewage Purification Degree Based on Fuzzy Causal Cluster Prediction Method

Yingying Jiang

School of Mathematics and Statistics, Suzhou University, Suzhou Anhui
Email: right067@126.com

Received: Dec. 28th, 2020; accepted: Jan. 26th, 2021; published: Feb. 2nd, 2021

Abstract

In order to study the law of the sewage purification degree treatment, this paper improves the fuzzy feature supply method of causal cluster analysis in the static fuzzy prediction. Typically, a new fuzzy feature set is established and the corresponding normal fuzzy number prediction model is proposed. In the empirical part, a causal compact class analysis based on the actual measured data is conducted to verify the model.

Keywords

Static Fuzzy Prediction, Causal Cluster Analysis, Fuzzy Feature Set, Normal Fuzzy Number

Copyright © 2021 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. 引言

自 1965 年 Zadeh [1] 定义了模糊集及运算法则, 开辟出了信息不确定的模糊数据处理的方法, 先后出现直觉模糊数[2]、三角模糊数、梯形模糊数[3]型等不同的模糊信息量化的形式, 并得到广泛研究。为避免隶属函数线性突变点本文选取正态模糊数[4], 利用正态模糊数的期望与方差刻画模糊决策值。方欢[5], 麻兴斌[6]分别对因果聚类分析进行研究, 本文采用模糊聚类与模糊模式识别的方法对静态模糊预测中模糊特征集的提取进行改进, 利用因素变量去推测预测值空间。

2. 备用知识

2.1. 正态模糊数

定义 1 [4]: $\tilde{A} = (a, \sigma^2)$ 为正态模糊数, 其中, $x, a \in R, \sigma > 0$ 。其隶属函数

$$\tilde{A}(x) = \exp\left(-\frac{(x-a)^2}{\sigma^2}\right).$$

正态模糊数运算法则:

设 $\tilde{A} = (a, \sigma_a^2), \tilde{B} = (b, \sigma_b^2)$, 则

1) $\tilde{A} + \tilde{B} = (a+b, \sigma_a^2 + \sigma_b^2)$; 2) $\lambda\tilde{A} = (\lambda a, \lambda^2 \sigma_a^2), \lambda \in [0, 1]$; 3) $\frac{1}{\tilde{A}} = \left(\frac{1}{a}, \frac{1}{\sigma_a^2}\right)$, 其中 $a \neq 0$ 。

2.2. 数据均值规格

特性指标矩阵 U 的第 j 列数据均值规格化[7], 计算

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (u_{ij} - \bar{u}_j)^2, \quad \bar{u}_j = \frac{1}{n} \sum_{i=1}^n u_{ij}, \quad j = 1, 2, \dots, m$$

作变换

$$u'_{ij} = \frac{u_{ij}}{\sigma_j}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m.$$

3. 静态模糊预测模型

模糊因果聚类

设有 T 期数据 $w_t = (x_t, v_t)$, ($t = 1, 2, \dots, T$), 其 $x_t = (x_{t1}, x_{t2}, \dots, x_{tm}) \in X_1 \times X_2 \times \dots \times X_m$

$v_t \in V$ 。 X 为因素状态空间, V 为预测空间。

数据矩阵为

$$W = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \end{pmatrix} = \begin{pmatrix} x_1 & v_1 \\ x_2 & v_2 \\ \vdots & \vdots \\ x_T & v_T \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} & v_1 \\ x_{21} & x_{22} & \cdots & x_{2n} & v_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{Tn} & v_T \end{pmatrix}$$

利用夹角余弦法 $r_{ij}(h) = \frac{|x_i \cdot x_j|}{\|x_i\| \cdot \|x_j\|}$, $\|x_i\| = \left(\sum_{k=1}^m x_{ik}^2\right)^{\frac{1}{2}}$, 求得模糊相似矩阵 $R(h) = (r_{ij}(h))_{T \times T}$, 利用多目标综合决策法求出最优的模糊相似矩阵 $R = (r_{ij})_{T \times T}$, 根据置信区间选出最佳聚类 U_1, U_2, \dots, U_m 。

4. 新模糊特征提取

将 U_i 向因素轴 X 投影, 得 $Y_i = \{x_i : w_i \in U_i\}$, $i = 1, 2, \dots, m$ 。对应 Y_i 构造原因特征模糊集 $A_i \in F(X)$ 。

$Y_i = \{x_{i1}, x_{i2}, \dots, x_{ik}\}$, 计算 Y_i 的几何均值:

$$\bar{x}_i = \frac{1}{k} \sum_{s=1}^k x_{is}, \quad \bar{x}_i = (\bar{x}_{i1}, \bar{x}_{i2}, \dots, \bar{x}_{in}), \quad (1)$$

$$\delta_{ij}^p = \frac{1}{k} \sum_{s=1}^k (x_{isj} - \bar{x}_{ij})^p, \quad j = 1, 2, \dots, n, \quad p \in N^+. \quad (2)$$

根据 \bar{x}_i 和 δ_{ij} 构造 Y_i 的模糊原因特征集 A_i 。

$$A_i(x) = \sum_{j=1}^n \alpha_j \exp\left\{-\frac{1}{4\delta_{ij}^2}(x_j - \bar{x}_{ij})^2\right\}, \quad (3)$$

其中 $(\alpha_1, \alpha_2, \dots, \alpha_n)$ 为一组给定的权重。

将 U_i 向预测轴 V 投影, 得 $Z_i = \{v_i : w_i \in U_i\}$, $i = 1, 2, \dots, m$ 。对应 Z_i 建立模糊数 $r_i \in \tilde{R}$ 。

$Z_i = \{v_{i1}, v_{i2}, \dots, v_{ik}\}$, 计算 Z_i 的几何均值:

$$\bar{v}_i = \frac{1}{k} \sum_{j=1}^k v_{ij}, \quad (4)$$

$$\delta_i^p = \frac{1}{k} \sum_{j=1}^k (v_{ij} - \bar{v}_i)^p. \quad (5)$$

(构造以 $(v_i, 2\delta_i)$ 为参数的正态模糊数 $r_i (i = 1, 2, \dots, m)$)。

综上得对应分类 $\{U_1, U_2, \dots, U_m\}$ 的特征模型为

$$\begin{pmatrix} U_1 & U_2 & \cdots & U_m \\ A_1 & A_2 & \cdots & A_m \\ r_1 & r_2 & \cdots & r_m \end{pmatrix}$$

进行预测若已知第 T 期的因素状态, 对 $\{U_1, U_2, \dots, U_m\}$ 和 x_t 应用 r_i 选出相应特征数对应正态模糊数作为 v 的预测值。

5. 实例分析

污水处理厂处理黑臭水, 选取以下五个因素预测污水净化程度 $C = \{c_1, c_2, c_3, c_4, c_5\}$, 其中 c_1 : 产量——每天污水处理量(千吨/天), c_2 : 质量——每天生化需氧量 BOD_5 去除率(%), c_3 : 质量——悬浮物 SS

去除率, c_4 : 气水比——处理每吨污水消耗的氧气的量(米³/千克), c_5 : 单耗——去除每千克 BOD_5 的耗电量(千瓦时/千克)。数据参见表 1。

Table 1. Information of the contamination of sewage purification
表 1. 污水净化处理数据信息

编号	c_1	c_2	c_3	c_4	c_5	污水净化程度 v
1	56	0.14	0.25	19	0.36	0.86
2	64	0.20	0.08	8	0.20	0.64
3	44	0.28	0.11	14	0.03	0.77
4	28	0.50	0.17	26	0.11	0.81
5	50	0.20	0.14	6	0.10	0.69
6	35	0.26	0.13	11	0.15	0.68
7	41	0.26	0.21	14	0.17	0.66

上述表中数据为 $w_t = (x_t, v_t)$, ($t = 1, 2, \dots, 7$), 其 $x_t = (x_{t1}, x_{t2}, \dots, x_{t5}) \in X_1 \times X_2 \times \dots \times X_5$
 v_t 是污水净化程度。

$$W = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \end{pmatrix} = \begin{pmatrix} x_1 & v_1 \\ x_2 & v_2 \\ \vdots & \vdots \\ x_T & v_T \end{pmatrix} = \begin{pmatrix} 56 & 0.14 & 0.25 & 19 & 0.36 & 0.86 \\ 64 & 0.20 & 0.08 & 8 & 0.20 & 0.64 \\ 44 & 0.28 & 0.11 & 14 & 0.03 & 0.77 \\ 28 & 0.50 & 0.17 & 26 & 0.11 & 0.81 \\ 50 & 0.20 & 0.14 & 6 & 0.10 & 0.69 \\ 35 & 0.26 & 0.13 & 11 & 0.15 & 0.68 \\ 41 & 0.26 & 0.21 & 14 & 0.17 & 0.66 \end{pmatrix}$$

对初始数据进行规格化[7]得矩阵 W' , $W' = \begin{pmatrix} 4.91 & 1.31 & 3.41 & 3.02 & 3.74 & 11.10 \\ 5.61 & 1.88 & 1.09 & 1.27 & 2.08 & 8.26 \\ 3.85 & 2.63 & 1.50 & 2.22 & 0.31 & 9.94 \\ 2.45 & 4.69 & 2.32 & 4.13 & 1.14 & 10.45 \\ 4.38 & 1.88 & 1.91 & 0.95 & 1.04 & 8.90 \\ 3.07 & 2.44 & 1.77 & 1.75 & 1.56 & 8.77 \\ 3.60 & 2.44 & 2.87 & 2.22 & 1.77 & 8.52 \end{pmatrix}$

利用多元分析方法来确定各对象间的相似度, 夹角余弦法 $r_{ij}(h) = \frac{|x_i \cdot x_j|}{\|x_i\| \cdot \|x_j\|}$, $\|x_i\| = \left(\sum_{k=1}^m x_{ik}^2 \right)^{\frac{1}{2}}$, 求得模

糊相似矩阵 $R(h) = (r_{ij}(h))_{T \times T}$:

$$R(h) = \begin{pmatrix} 1 & 0.96 & 0.95 & 0.92 & 0.97 & 0.97 & 0.98 \\ 0.96 & 1 & 0.96 & 0.89 & 0.98 & 0.96 & 0.96 \\ 0.95 & 0.96 & 1 & 0.97 & 0.98 & 0.99 & 0.98 \\ 0.92 & 0.89 & 0.97 & 1 & 0.93 & 0.97 & 0.97 \\ 0.97 & 0.98 & 0.98 & 0.93 & 1 & 0.99 & 0.98 \\ 0.97 & 0.96 & 0.99 & 0.97 & 0.99 & 1 & 0.99 \\ 0.98 & 0.96 & 0.98 & 0.97 & 0.98 & 0.99 & 1 \end{pmatrix}$$

根据置信区间 $[0.89,1]$ ， $\lambda = 0.97$ 选出最佳聚类 U_1, U_2, U_3, U_4 的分类。

$$U_1 = \{I_2, I_3\}, U_2 = \{I_5, I_7\}, U_3 = \{I_1, I_6\}, U_4 = \{I_4\}$$

求分类 U_1, U_2, U_3, U_4 中各属性状态特征值的平均值，得

$$\bar{x}_1 = (4.73, 2.26, 1.30, 1.75, 1.20); \bar{x}_2 = (3.99, 2.16, 2.39, 1.59, 1.41)$$

$$\bar{x}_3 = (3.99, 1.88, 2.59, 2.34, 2.65); \bar{x}_4 = (2.45, 4.69, 2.32, 4.13, 1.14)$$

并求这4个分类的属性状态特征值的均值方差，得

$$\sigma_{11}^2 = 0.77, \sigma_{12}^2 = 0.14, \sigma_{13}^2 = 0.04, \sigma_{14}^2 = 0.23, \sigma_{15}^2 = 0.78;$$

$$\sigma_{21}^2 = 0.15, \sigma_{22}^2 = 0.08, \sigma_{23}^2 = 0.23, \sigma_{24}^2 = 0.40, \sigma_{25}^2 = 0.13;$$

$$\sigma_{31}^2 = 0.85, \sigma_{32}^2 = 0.32, \sigma_{33}^2 = 0.67, \sigma_{34}^2 = 0.41, \sigma_{35}^2 = 1.19;$$

$$\sigma_{41}^2 = 0.00, \sigma_{42}^2 = 0.00, \sigma_{43}^2 = 0.00, \sigma_{44}^2 = 0.00, \sigma_{45}^2 = 0.00.$$

带入实测污水净化程度接近4个分类中污水净化程度的平均值的 I_2, I_7, I_1, I_4 ，取权重 $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.1, 0.2, 0.35, 0.2, 0.15)$ ，带入(3)式求得各分类的特征模糊集为：

$$A_1(x) = \sum_{j=1}^5 \alpha_j \exp \left\{ -\frac{1}{4\delta_{ij}^2} (x_j - \bar{x}_{ij})^2 \right\},$$

$$= 0.1e^{-\frac{(x_1-4.73)^2}{3.08}} + 0.2e^{-\frac{(x_2-2.26)^2}{0.56}} + 0.35e^{-\frac{(x_3-1.30)^2}{0.12}} + 0.2e^{-\frac{(x_4-1.75)^2}{0.92}} + 0.15e^{-\frac{(x_5-1.20)^2}{3.12}},$$

$$A_2(x) = 0.1e^{-\frac{(x_1-3.99)^2}{0.6}} + 0.2e^{-\frac{(x_2-2.16)^2}{0.32}} + 0.35e^{-\frac{(x_3-2.39)^2}{0.92}} + 0.2e^{-\frac{(x_4-1.59)^2}{1.6}} + 0.15e^{-\frac{(x_5-1.41)^2}{0.52}},$$

$$A_3(x) = 0.1e^{-\frac{(x_1-3.99)^2}{3.4}} + 0.2e^{-\frac{(x_2-1.88)^2}{1.28}} + 0.35e^{-\frac{(x_3-2.59)^2}{2.68}} + 0.2e^{-\frac{(x_4-2.34)^2}{1.64}} + 0.15e^{-\frac{(x_5-2.65)^2}{4.76}},$$

$$A_4(x) = 0.00.$$

根据(4~5)式取 $p = 2$ 求污水净化程度的平均值，得

$$\bar{v}_1 = 9.1, \bar{v}_2 = 8.71, \bar{v}_3 = 9.94, \bar{v}_4 = 10.45;$$

$$\delta_i^2 = \frac{1}{k} \sum_{j=1}^k (v_{t_j} - v_i)^2, \text{ 则 } \delta_1 = 0.84, \delta_2 = 0.19, \delta_3 = 1.16, \delta_4 = 0.00.$$

构造正态模糊数为 $r_1 = (9.1, 1.68)$ ， $r_2 = (8.71, 0.38)$ ， $r_3 = (9.94, 2.32)$ ， $r_4 = (10.45, 0.00)$ 。

取 I_2, I_7, I_1, I_4 实测数据获得4个模糊特征集为

$$A_1(x) = 0.7643, A_2(x) = 0.7796, A_3(x) = 0.7732, A_4(x) = 0.00$$

由最大隶属度原则，应选择 r_2 作为模糊预测，与实测值规范化处理后8.52较接近。类似比较其他3个编号预测值与实际值其误差范围在0.19~1.16，误差范围满足要求，此预测模型可行。

6. 结论

基于因果聚类模糊预测采用夹角余弦法对数据进行回归分析，改进模糊特征集的提取方式获得模糊特征集规避了直接映射强线性；污水净化处理中选取的因素具有可控性，通过对预测目标影响因素的选择可对预测目标空间值进行控制，本文通过实例验证了静态模糊预测模型具有可行性，预测得目标值与实际值误差在允许范围内。

基金项目

安徽省高校自然科学重点研究项目(KJ2020A0731), 安徽省自然科学重大教学改革研究项目(2020zdxsjg390), 安徽省自然科学基金项目(2020szsfkc0998)。

参考文献

- [1] Zadeh, L.A. (1965) Fuzzy Sets. *Information and Control*, **8**, 338-353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- [2] Atanassov, K. (1986) Intuitionistic Fuzzy Sets. *Fuzzy Sets and Systems*, **20**, 87-96. [https://doi.org/10.1016/S0165-0114\(86\)80034-3](https://doi.org/10.1016/S0165-0114(86)80034-3)
- [3] 李荣均. 模糊多属性决策理论及其应用[M]. 北京: 科学出版社, 2002.
- [4] 彭祖赠, 孙韞玉. 模糊数学及其应用[M]. 武汉: 武汉大学出版社, 2007.
- [5] 方欢, 金朋朋, 方贤文, 等. 基于因果行为轮廓的流程变体聚类挖掘方法[J]. 计算机集成制造系统, 2020, 26(6): 1538-1547.
- [6] 麻兴斌, 孟祥君, 刁柏青. 基于因果聚类分析理论的维修成本控制体系研究[J]. 科技和产业, 2018, 18(6): 96-100.
- [7] 陈水利, 李敬功, 王向公. 模糊集理论及其应用[M]. 北京: 科学出版社, 2005.