

OPTIMIZED SELECTION OF WETLAND WATER QUALITY MONITORING POINTS BASED ON INFORMATION ENTROPY AND FUZZY SIMILARITY

Xinjian Xiang

College of Automation & Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou, 310023, P. R. China, Tel: +86-571-85070268, Email: hzxxj@sina.com

Abstract: Known as the kidney of earth, wetland has significant ecological functions such as freshwater conservation, poison elimination, carbon storage, water quality purification, flood storage and drought control, climate regulation and remaining biodiversity etc. So protecting wetland is protecting ourselves. Water environment quality best reflects the ecological environment condition of wetland. According to multi-index and Spatial and Temporal variation of wetland water pollution, combining optimized selection requirements of wetland water quality monitoring, fuzzy similarity is propose. Through constructing multi-index monitoring data samples Decision-making Matrix, fuzzy similarity matrix between sample data and their mean values is established. According to the index value variation, the index weights are calculated based on information entropy theory. With the index weight and sample fuzzy similarity matrix, comprehensive fuzzy similarity of each monitoring point is calculated. Finally, according to comprehensive fuzzy similarity, each monitoring point is reasonably clustered, then representative points is selected from each category, so distribution optimization could be realized. Practical running proves that this scheme is simple and feasible, and extensionally applied to optimize other environmental monitoring points.

Keywords: fuzzy similarity method, information entropy, wetland water quality monitoring, optimized selection

1. INTRODUCTION

Among three biggest ecosystems, known as the Kidney of Earth, wetland has significant ecological functions such as freshwater conservation, poison elimination, carbon storage, water quality purification, flood storage and drought control, climate regulation and remaining biodiversity etc. It plays an important role in remaining ecological balance and prosperous economy. Wetland ecological environment mainly includes water, air, acoustic and soil environment. And water quality best reflects wetland ecological environment condition (Liang et al., 2002). Protecting wetland, and establishing water environment real-time monitoring system is an urgent task.

Compared with sea, rivers and lakes, wetland water environment has its specialties (Jiang, 2007): entire water environment is divided into widely distributed small independent water areas with irregular shapes, different area. Area and depth of water area is easily influenced by season, climate, human interruption and other factors. As the obvious difference, it is necessary to monitor the water environment of each small water area in wetland with multi-points real-time monitoring. In the initial period of wetland water quality monitoring, generally there are a lot of points distributed to fully master the water quality of entire area. After a certain period of monitoring, abundant data are accumulated. To reduce Human resource, material resource and financial resource and other waste, optimized distribution is necessary. Optimized goal of distribution is to search minimum monitoring points and still objectively reflect area environment quality. Currently, there are many optimized methods applied to environment monitoring points, such as Fuzzy Clustering, Matter element and Osculation Value methods. Though with different characteristics, these methods all have complicated programs. Through constructing multi-index monitoring data samples Decision-making Matrix, fuzzy similarity matrix between sample data and their mean values is established. According to the index value variation, the index weights are calculated based on information entropy theory. With the index weight and sample fuzzy similarity matrix, comprehensive fuzzy similarity of each monitoring point is calculated. Finally, according to comprehensive fuzzy similarity, each monitoring point is reasonably clustered, then representative points is selected from each category, so distribution optimization could be realized. Compared with other methods, fuzzy similarity has clear conception, simple calculation, direct image and single conclusion. Practical running proves that after analysis, the results of optimized points are corrects.

2. FUZZY SIMILARITY MONITORING POINTS OPTIMIZATION THEORY

Wetland water quality environment changes with space and time (Xu et al., 2002). In addition, multi-index brings lots of difficulties to monitoring point's optimization. Single index parameter comprehensively reflects water quality. As a result, how to transform multi-index monitoring parameters into single index parameter and then cluster monitoring points is the basic access to optimize monitoring points and the starting point of proposing and applying the concept of fuzzy similarity.

2.1 Sample Decision Making Matrix Establishment

m monitoring points ($\theta_1, \theta_2, \dots, \theta_m$) n evaluation index ($A_1, A_2 \dots A_n$), initial matrix R_0 is constituted:

$$\begin{matrix}
 A_1 & A_2 & \dots & A_n \\
 \theta_1 & \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{bmatrix} \\
 \theta_2 & \\
 \vdots & \\
 \theta_m &
 \end{matrix} = (c_{ij})_{m \times n} \tag{1}$$

C_{ij} is the i th monitoring point θ_i 's j th quantized value of evaluation index, $i \in (1, 2 \dots m), j \in (1, 2 \dots n)$.

$$r_{ij} = \frac{C_{ij}}{\sqrt{\sum_{i=1}^m C_{ij}^2}} \tag{2}$$

Normalize Matrix (1) establish sample decision making matrix R :

$$\begin{matrix}
 A_1 & A_2 & \dots & A_n \\
 \theta_1 & \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix} \\
 \theta_2 & \\
 \vdots & \\
 \theta_m &
 \end{matrix} = (r_{ij})_{m \times n} \tag{3}$$

$\theta_i = (r_{i1}, r_{i2}, \dots, r_{in})$ ($i = 1, 2 \dots m$) is the sample point to be optimized.

2.2 Fuzzy Similarity Theory

2.2.1 Fuzzy process of measurement value and mean value

Influenced by various factors, there is error between water quality index sensors' measurement value and object real value. Real value is near all the effective values. Supposing measurement value error is random error, measuring distribution is normal distribution decided by mean value and variance. Fuzzy membership function takes Gauss. To satisfy engineering application needs, triangle membership function is adopted here. The center of triangle is sensor's normalized measurement value; the width is four times of measurement data standard various. For sensor i , suppose measuring mean value after L times measurements value of real value A of certain monitoring point is x_i , measuring variance is σ_i and measuring value fuzzy value can be expressed as follows:

$$\tilde{A}_i = (a_{i1}, a_{i2}, a_{i3}) = (x_i - 2\sigma_i, x_i, x_i + 2\sigma_i)$$

Mean value (normalized) fuzzy processing is similar with that of measuring value. Only mean value x_0 , mean variance σ_0 are as follows (m is monitoring points number):

$$x_0 = \frac{1}{m} \sum_{i=1}^n x_i$$

$$\sigma_0^2 = \frac{1}{m-1} \sum_{i=1}^n (x_i - x_0)^2 \tag{4}$$

S_0 fuzzy value of mean value is:

$$\tilde{A}_0 = (a_{01}, a_{02}, a_{03}) = (x_0 - 2\sigma_0, x_0, x_0 + 2\sigma_0)$$

2.2.2 Definition and calculation of fuzzy similarity

Suppose \tilde{A}_i and \tilde{A}_j are two fuzzy values, defining $S = S(\tilde{A}_i, \tilde{A}_j)$, if S satisfies:

- (1) $0 \leq S \leq 1$.
- (2) for $\tilde{A}_i = \tilde{A}_j$, $S = 1$.
- (3) $S(\tilde{A}_i, \tilde{A}_j) = S(\tilde{A}_j, \tilde{A}_i)$.

(4) If and only if $\tilde{A}_i \cap \tilde{A}_j = \Phi$, $S(\tilde{A}_i, \tilde{A}_j) = 0$.

(5) if $\tilde{A}_i \subset \tilde{A}_j \subset \tilde{A}_s$, $S(\tilde{A}_i, \tilde{A}_j) > S(\tilde{A}_i, \tilde{A}_s)$.

Then call S as \tilde{A}_i, \tilde{A}_j similarity, that is close extension of \tilde{A}_i and \tilde{A}_j .

There are many methods for calculating similarity between fuzzy values (Xiang, 2004). To realize reliability and convenience, define similarity calculation method based on distance measurement:

Suppose $\tilde{A}_i = (a_{i1}, a_{i2}, a_{i3})$ and $\tilde{A}_j = (a_{j1}, a_{j2}, a_{j3})$ are two triangle fuzzy value, then the similarity is:

$$S(\tilde{A}_i, \tilde{A}_j) = \frac{1}{1 + d(\tilde{A}_i, \tilde{A}_j)} \tag{5}$$

In the equation:

$$d(\tilde{A}_i, \tilde{A}_j) = \frac{a_{i1} + 4a_{i2} + a_{i3} - a_{j1} - 4a_{j2} - a_{j3}}{6}$$

The bigger $S(\tilde{A}_i, \tilde{A}_j)$ is, the closer \tilde{A}_i and \tilde{A}_j . $S(\tilde{A}_i, \tilde{A}_j) = 1$, means \tilde{A}_i and \tilde{A}_j are completely the same, $S(\tilde{A}_i, \tilde{A}_j) = 0$, means \tilde{A}_i and \tilde{A}_j totally different. To convenience describe, define \tilde{A}_{ij} as the normalized fuzzy measurement value of j , index of point i , \tilde{A}_{0j} 为 is fuzzy value of normalized measurement mean value, so $S_{ij} = S(\tilde{A}_{ij}, \tilde{A}_{0j})$ is called as fuzzy similarity of index j of monitoring point i .

3. OPTIMIZED SELECTION OF WETLAND WATER QUALITY MONITORING POINTS BASED ON INFORMATION ENTROPY AND FUZZY SIMILARITY METHOD

Information entropy and fuzzy similarity optimization principle is a level entropy multi-object comprehensive evaluation method. Based on

information entropy and fuzzy similarity, basic process of optimizing water quality monitoring points is: construct multi-index parameter sample matrix of monitoring points, normalize the matrix; establish sample data and the fuzzy similarity matrix between ample data and their mean value, with information entropy technology, each index weight is confirmed. According to this, comprehensive similarity of monitoring and mean value points is calculated. And sequencing cluster the monitoring points by similarity, finally realize optimized distribution according to clustering results.

3.1 Index Weight Confirmation Principles and Methods Based Information Entropy

In multi-objects decision making, the bigger the difference of certain index value, the more important it is in comprehensive evaluation. If index values of certain index are all the same, the index is useless in comprehensive evaluation. In information theory, entropy means information quantity got from a group of unknown objects. Information entropy is the measurement of disorder extension of a system (Sun et al., 2000). The bigger the variance extension of a certain index, the smaller the information entropy is, that is the function is much stronger in comprehensive evaluation. So the index weight is bigger; in the contrary, the weight should be smaller. According to the variance of each index, with information entropy, weight of each index can be calculated, specific method as follows:

- (1) Calculate output entropy of index j:

$$E_j = -(\ln m)^{-1} \sum_{i=1}^m p_{ij} \ln p_{ij} \tag{6}$$

Of which: $p_{ij} = r_{ij} / \sum_{i=1}^m r_{ij}$

If $p_{ij}=0$, then define $p_{ij} \ln p_{ij}=0$.

- (2) Calculate variety of index j

$$D_j = 1 - E_j \tag{7}$$

- (3) Calculate objective weight of index j

$$w_j = D_j / \sum_{j=1}^n D_j \tag{8}$$

3.2 Fuzzy Similarity Optimized Selection of Wetland Water Quality Monitoring Points

As shown above, in the foundation of getting the fuzzy similarity of the index j of point i, and calculating relative weight of each index,

comprehensive fuzzy similarity of point i can be calculated with the following equation:

$$S_i = \sum_{j=1}^n W_j S_{ij} \quad (9)$$

According to above thoughts, the optimized selection of wetland water quality monitoring distribution based on information entropy and fuzzy similarity is as follows:

- 1) With equations (1), (2), (3), n index evaluation data of m monitoring points are normalized and sample matrix R is established.
- 2) With equation (4), (5), the index j of point i fuzzy similarity is calculated. Fuzzy similarity matrix of sample data and their mean value is established.
- 3) With equation (6), (7), (8), relative weight in index j in clustering data is confirmed.
- 4) With equation (9), the comprehensive fuzzy similarity of point i is calculated finally.
- 5) According to the comprehensive fuzzy similarity, points are scientifically clustered. And representative points are selected from each category to realize distribution optimization.

4. EXAMPLE ANALYSIS

With information entropy and fuzzy similarity mentioned above, provided by a certain environment monitoring station in a city's wetland in 2003, data of 7 monitoring points have been optimized. Table 1 proves that, there are 4 pollution factors, respectively as COD, BOD₅, NH₃-N and TP.

Table 1. The water quality monitoring data

Moni-points	COD	BOD ₅	NH ₃ -N	TP
1	14	3.1	1.56	0.138
2	24	2.8	2.06	0.134
3	18	4.2	0.31	0.037
4	33	2.9	0.99	0.099
5	34	5.9	1.85	0.385
6	19	4.8	2.60	0.486
7	17	5.1	1.08	0.265

According to methods mentioned above, we first establish sample matrix $(C_{ij})_{7 \times 4}$ of 7 monitoring points data, then normalize the matrix to gain dimensionless matrix $(r_{ij})_{7 \times 4}$. Applying equation (4) and (5), fuzzy similarity S_{ij} is calculated, shown as table 2.

Table 2. The similarity degree between the monitoring points and the standard-point: Ci-m

S_{ij}	1	2	3	4
1	0.3799	0.2481	0.4713	0.1889
2	0.2849	0.3721	0.0709	0.0522
3	0.5223	0.2569	0.2265	0.1396
4	0.5382	0.5228	0.4233	0.5429
5	0.3007	0.4253	0.5949	0.6853
6	0.2691	0.4519	0.2471	0.3737
7	0.3595	0.3645	0.3415	0.3111

According to (6), (7), (8) relative weight of index j in data clustering is confirmed, shown as table 3.

Table 3. The relative weight of index j

Index	1	2	3	4
w_j	0.188	0.167	0.297	0.348

From equation (9) fuzzy similarity of point i is finally calculated, shown as table 4.

Table 4. The similarity degree between the monitoring points and the average values

Points	1	2	3	4	5	6	7
S_i	1.09	1.03	1.24	1.10	0.86	0.85	1.02

Seen from table 4: we can cluster points 1 and 4 as the same category; 2 and 7, 5 and 6. According to clustering results, best monitoring points can be selected: points 3, 1, 7 and 5, of which point 5 is the representative point of most serious pollution, and 7 secondary, 1 common, and point 3 is the list.

5. CONCLUSIONS

Monitoring points optimization based on information entropy and fuzzy similarity is a multi-objects decision-making method. It can cluster and sort. Applying level analysis, entropy technology and fuzzy similarity principles, optimized decision-making mode is established, which effectively provides scientific decision-making basis for setting wetland water quality monitoring points.

Practical examples prove that applying this method to optimize distribution of wetland water quality monitoring points, optimized results are acceptable. Information entropy and fuzzy similarity have clear calculation significance and concept, flexible and convenient, which can be extended to optimized distribution in other environments. Meanwhile, seen from the concept of fuzzy similarity concept, it fully mines the abundant information in the sample data and could be called as a good data processing method.

REFERENCES

- Jiang Peng, Survey on key technology of WSN-based wetland water quality remote real time monitoring system [J]. Chinese journal of sensors and actuators No. 1, 2007, pp. 83-86.
- Liang weizhen, ye jinglun. Water quality evaluation based on optimal fuzzy clustering [J]. The Administration and Technique of Environmental Monitoring No. 6, 2002, pp. 6-7.
- Sun Shimin, Shi Haixing, Evaluation of Milking Machine Based on Entropy Technology and Idea 1 Point Principle [J]. Transactions of the Chinese Society for Agricultural Machinery, No. 5, 2007, pp. 82-87.
- XiangXinjian, A method to sensor data fusion based on fuzzy and statistics integration Chinese journal of sensors and actuators, No. 2, 2004, pp. 197-199.
- Xu lizhong, Zhang jiangshang. Optimization of atmospheric environmental monitoring sites with modified intimate value method [J]. Environmental Engineering No. 18, 2002, pp. 50-53.