*Desalination and Water Treatment*
www.deswater.com

doi: 10.5004/dwt.2019.24398

167 (2019) 1–12
November

# Management of algae bloom based on CBR-OSS model

Zhao-Yang Wang[a,b], Bai-Hai Zhang[a,*], Xiao-Yi Wang[b], Hui-Yan Zhang[b], Ji-Ping Xu[b], Yu-Ting Bai[a]

[a]*School of Automation, Beijing Institute of Technology, Beijing 100081, China, emails: smczhang@bit.edu.cn (B.-H. Zhang), sophiewong0121@126.com (Z.-Y. Wang), byting@bit.edu.cn (Y.-T. Bai)*
[b]*School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China, emails: wangxy@btbu.edu.cn (X.-Y. Wang), zhanghy@btbu.edu.cn (H.-Y. Zhang), xujp@th.btbu.edu.cn (J.-P. Xu)*

Received 8 September 2018; Accepted 10 May 2019

**ABSTRACT**

The outbreak process of algal bloom is a complex ecological problem of system engineering involving various factors such as water parameters, surrounding environment and human activity. For this ecological problem, the strict restriction and requirement limit the development of management about algae bloom. To select the most suitable strategy from various algae control methods, we propose case-based reasoning-optimal strategy selection (CBR-OSS) model. It builds case library and complex network by extracting the factors of algae management. This model regards the complex network as a directive network to reflect dynamic characteristic and weights of key factors. To improve decision efficiency, it defines the restriction slots and condition slots in directive network. As the inference engine, these slots exclude the unsuitable cases and avoid the redundancy computation so that the model can calculate the similarity between the target water body and screen cases in the process of decision case matcher. This process finds the best matching case and recommended measures by intuitionistic fuzzy rough sets. To verify the model, Kunming Lake and other 20 lakes are simulated with the proposed method. The results accord with expert advice and the model outperforms in accuracy, operation time, expert participation and flexibility.

*Keywords:* Algae bloom; Management strategy; Case-based reasoning; Optimal strategy selection

## 1. Introduction

Lake eutrophication is a water pollution phenomenon caused by extra increase of nutrient salt (including nitrogen and phosphorus) which increases the productivity (or the rate of photosynthesis) of the aquatic ecosystem [1]. Negatively, the surplus productivity promotes the growth of harmful algae blooms (HABs). This phenomenon and its damage have emerged all over the world. As shown in report, HABs have involved places such as East-Asia [2], Africa [3], Europe [4] and North America [5] where nearly 70% lakes and reservoirs statistically suffer from eutrophication and HABs [6]. The overgrowing algae have severe effect on water landscape, aquaculture and the safety of drinking water. Its threat has directed attention to research of HAB management in terms of the protection and recovery of water environment [7,8].

Macroscopically management strategies of algae blooms are categorized into biological, chemical and physical methods [9]. Among them, the most common method is physical method. One physical way to reduce the blooms is shock effect by smashing and coagulating blooms [10,11]; another physical way to reduce the blooms is the advanced filter membrane with physical filtration. It accomplishes the removal by adsorbing the algae and magnetite for a given

* Corresponding author.

residence time in a magnetic filter [12,13]. Sludge dredging is also a physical way to prevent and control algae blooms by reducing nutrient N and P [14]. The management result of physical methods is obvious and quick, but the expense is too high to be suitable in large water body. To avoid the disadvantages of physical method, chemical method become a popular method. Chemical method includes chemical deposition and strong oxidants method. The strong oxidants method aims to destroy the algae cells by adding strong oxidants into water [15,16]. And chemical deposition adsorbs and coagulates algae cells by coagulants [17,18]. Compared with physical methods, the cost of chemical methods is acceptable. However the secondary pollution is the most serious problem which limits the application of chemical methods. Considering problems of physical and chemical methods, biological methods are proposed in directions of biological nutrient control and biological algae inhibition. Biological algae removal suppresses the algae growth by bacteria and plants. These interactions between bacteria and species are potentially important factors affecting both the population dynamics of algae and its toxicity [19,20]. For biological nutrient control, some bacteria and plants reduce the nutrient density and suppress the algae growth by effective nutrient removal [21]. Biological methods above advance in the implementation process and result, but the ecological safety problem should not be ignored.

In general, each method above cannot meet the requirements for all lakes and reservoirs in terms of their advantages and disadvantages. To find a widely-used management strategy, some researchers engage in optimal strategy selection (OSS) of algal management. They explore to select the best management strategy from biological, chemical and physical methods for a specific water body. This way does not avoid negative effect of the method, but the best management strategy meets the special requirements for the specific water body. Meanwhile it is easier to realize in software system and outperform in information analysis. By gathering the lakes' information of water and environment, researchers analyze the requirements and process the information by artificial intelligent (AI) or other methods to obtain optimal strategy. The first intelligent method applied in algae control is multiple attribute decision making (MADM) [22]. Based on mechanism characteristic of algae growth, MADM substitutes information process technology for expert advice. Considering fuzzy problem, multi-objective method combined with vague set theory is also introduced to better select the management strategy [23]. However these decision-making processes fully depends on subjective information and only reflects the general features of the methods which ignore the relevance to the real-time environmental status. To solve the problem, researchers proposed a novel group decision-making method fused with sensor information [24,25]. In practicality and applicability, OSS has a better performance compared with single management strategy. However, some of the OSS methods have not been tested strictly while they only remain at experiment or simulation. If these methods are applied in water body, the outcome leaves us worried about secondary damage of water body.

Since OSS is a complex ecological problem, it may involve various factors in implementation process with only water quality information considered before. But these factors also include surrounding environment, government expenditure, climate, etc. Existing research does not fully consider the intrinsic properties of water body and HABs. Impacted on growth mechanism of HABs, there existed interaction among factors [26]. This interaction would lead to the uncertainty and randomness in management process so that the decision may be not complete and precise. Based on the research above, we proposed CBR-OSS model to select and inherit good management cases. The CBR-OSS model ranks the existed management methods and selects the most suitable method which applies to the specific water body. To build the CBR-OSS model, cases of algae management having been applied in lakes and reservoirs worldwide before are first gathered in case library. Then we analyze and extract key factors of algae management as the nodes of network which reflect the intrinsic properties of algae management. The linkage of inner factors forms a dynamic complex network in which the weights of factors are given. To find the best strategy quickly, the screening by inference rules is designed to exclude dissimilar cases in case library. For the similar cases, intuitionistic fuzzy rough sets (IFRS) calculate the similarity between the target case and cases in screening case library. By comparing the similarity degree, the possible strategies are ranked and selected.

The remainder of this paper is organized as follows. Section II presents the system model. Section III introduces our CBR-OSS method. Under the proposed method, Section IV presents the experiment and result. Section V concludes this paper. We also give an abbreviation list (Appendix I).

## 2. System model

The optimal selection aims to find the best management strategy with accessible information of water body. When key factors are regarded as the known variables, they are processed mathematically or logically to determine final decision. This process can be abstracted into a decision function based on CBR theory. CBR is an approach which solves a new problem by retrieving past cases and reusing their solutions. Because this process reasons and operates in a way of human beings, it is an important method in the fields of artificial intelligence (AI). CBR helps decision-makers to find the desirable solutions to decision-making problems. It usually includes five steps as shown in Fig. 1, that is, case presentation, case retrieval, case reusing, case revision, case retaining.

The operation process of OSS is shown in Fig. 2. This process corresponds to framework of CBR. Case presentation, the first step, combines the key factors extraction and network modeling. The key factors constitute the dynamic complex network by defining the importance degree function which optimizes the key factor weight. Then CBR-OSS model operates the inference engine and decision matcher automatically. These operations screen cases to get optimal strategy, but the optimal strategy is only reused after being examined by domain expert. Finally, the successful application is retained in case library so that self-learning is completed.

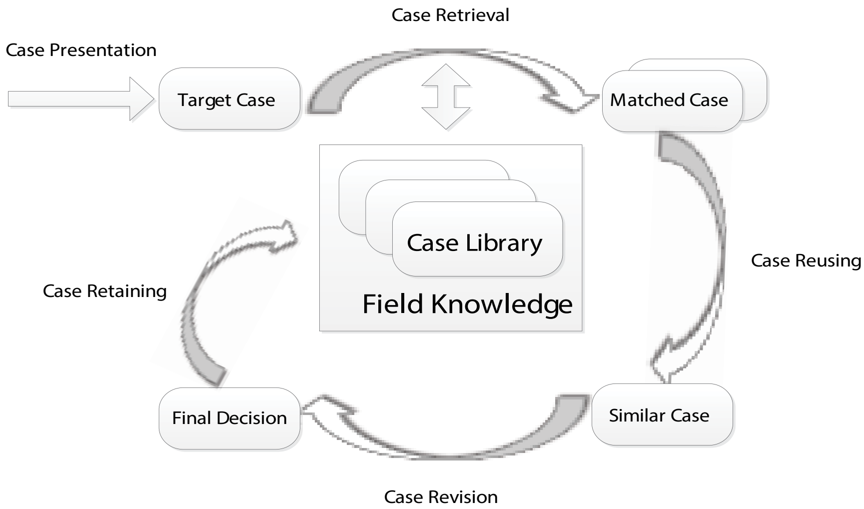


Fig. 1. Process of CBR.

## 3. Methods

### 3.1. Complex network of dynamic relation

To better describe the decision process and find the most suitable strategy, the selection of key factors in algae management is of great significance. After combining the past research and expert opinion, the key factors are macroscopically categorized into variables $V$ = {outbreak scene of algae, water parameters, economics, cultural environment, natural environment} which are defined as the nodes of total network. Each node in total network can be divided into a sub-network. The factors in sub-network are shown as follows. For outbreak scene of algae, $V_1$ = {surface color, algae species, eutrophication level, algae area, surface properties, smell}. For water parameters, $V_2$ = {total phosphorus (TP), total nitrogen (TN), ratio of nitrogen and phosphorus (N/P), pH, electronic conductivity (EC), temperature of water (TW), dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), Chl a}. For economics, $V_3$ = {response time, investment of water treatment, long-term governance, ecological security, second pollution}. For cultural environment, $V_4$ = {domestic sewage discharge, utilization of water, population mobility, agricultural wastewater discharge, industrial wastewater discharge}. For natural environment, $V_5$ = {area of water, lake type, surrounding environment, air temperature, humidity, light intensity, geographical position, wind speed}.

Obviously, the dynamic relation and interaction exist among key factors. It means the complex network can be built by this dynamic relation. If the interaction exists between two factors, the complex network adds an edge for these two nodes. On the basis, the dynamic complex network is designed as shown in Fig. 3.

Complex network is an abstraction method of complex system which associates the key factors of algae management. In complex network, the node importance reflects the weight of key factor, and the edge describes the interaction among key factors [27]. Let $G(V,E)$ denote a complex network, where $V = \{v_1,v_2,\cdots,v_i,\cdots,v_n\}$ and $E = \{e_1,e_2,\cdots,e_j,\cdots,e_m\}$ denote the node set and edge set of the network, respectively. $n$ and $m$ are the numbers of nodes and edges. If $v_i$ and $v_j$ are connected by an edge, element of the network adjacency matrix is $\delta_{ij} = 1$, otherwise $\delta_{ij} = 0$. The network adjacency matrix is also called contribution allocation parameter. For each node, characteristic parameters degree $\lambda_i$, in-degree $\lambda_i^+$, out-degree $\lambda_i^-$ are defined where $\lambda_i = \sum_{j=1}^{n}\delta_{ij}, i = 1,2,\cdots,n$. In-degree and out-degree are in directive network. $d_{ij}$ is the node distance which denotes the shortest distance from node $v_i$ to $v_j$. Therefore, network efficiency $E$ and the node efficiency $I_k$ are given by Eqs. (1) and (2).

$$E = \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{d_{ij}} \tag{1}$$

$$I_k = \frac{1}{n}\sum_{i=1,i\neq k}^{n}\frac{1}{d_{ki}} \tag{2}$$

By analyzing the dynamic relation of complex network, the importance of key factors is easily obtained. Both the network efficiency and the node efficiency affect the node weight. In unidirectional network, the in-degree and out-degree have no difference for importance evaluation Eq. (3). But it is not suitable for the complex network of algae management.

$$H_{\mathrm{IC}} = \begin{bmatrix} 1 & \delta_{12}\lambda_2/\langle k\rangle^2 & \cdots & \delta_{1n}\lambda_n/\langle k\rangle^2 \\ \delta_{21}\lambda_1/\langle k\rangle^2 & 1 & \cdots & \delta_{2n}\lambda_n/\langle k\rangle^2 \\ \vdots & \vdots & \cdots & \vdots \\ \delta_{n1}\lambda_1/\langle k\rangle^2 & \delta_{n2}\lambda_2/\langle k\rangle^2 & \cdots & 1 \end{bmatrix} \tag{3}$$

The interaction among key factors in algae network is directive. To reflect the direction in complex network, we propose a new importance formula in which $\lambda_i/\langle k\rangle^2$ is replaced by $\left(\lambda_i^{\lambda_i^-/\lambda_i}\right)^2/\langle k\rangle^2$. The improved importance evaluation formula is shown as follows:

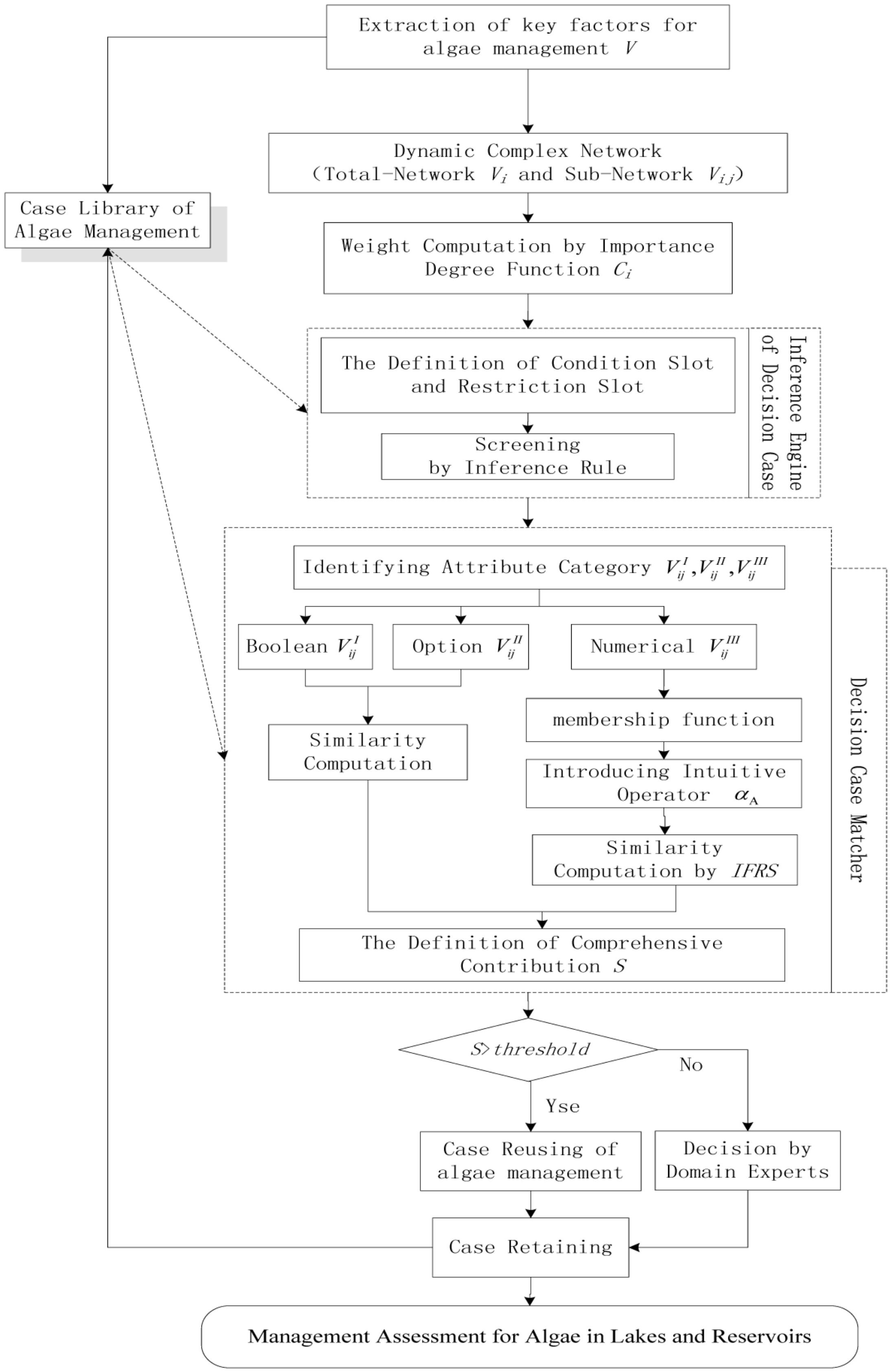


Fig. 2. Working flow of CBR-OSS model.

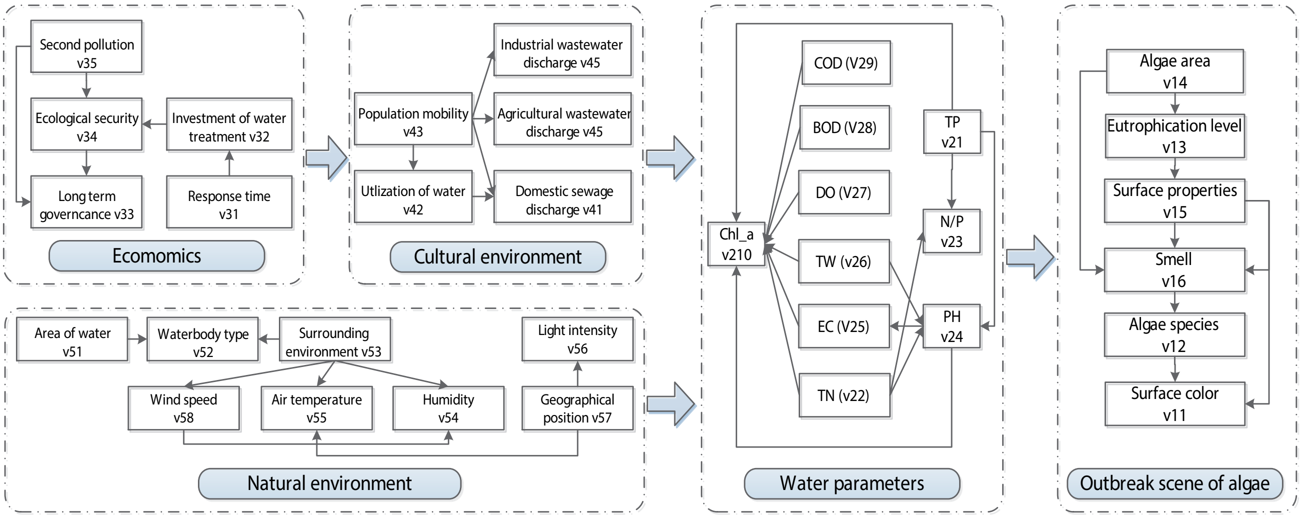


Fig. 3. Dynamic complex network of key factors.

$$H_E = \begin{bmatrix} I_1 & \delta_{12}\left(\lambda_2^{\lambda_2^-/\lambda_2}\right)^2 I_2/\langle k\rangle^2 & \cdots & \delta_{1n}\left(\lambda_n^{\lambda_n^-/\lambda_n}\right)^2 I_n/\langle k\rangle^2 \\ \delta_{21}\left(\lambda_1^{\lambda_1^-/\lambda_1}\right)^2 I_1/\langle k\rangle^2 & I_2 & \cdots & \delta_{2n}\left(\lambda_n^{\lambda_n^-/\lambda_n}\right)^2 I_n/\langle k\rangle^2 \\ \vdots & \vdots & \cdots & \vdots \\ \delta_{n1}\left(\lambda_1^{\lambda_1^-/\lambda_1}\right)^2 I_1/\langle k\rangle^2 & \delta_{n2}\left(\lambda_2^{\lambda_2^-/\lambda_2}\right)^2 I_2/\langle k\rangle^2 & \cdots & I_n \end{bmatrix} \quad (4)$$

where $H_E$ is importance matrix. $H_{Eij}$ denotes the importance value from node $j$ to node $i$. Synthesizing local importance and global importance of nodes, the importance degree is defined as the combination of nodes and their adjacent nodes. The improved importance formula is shown in Eq. (5) and it is used to denote weights of network nodes.

$$C_i = I_i \times \sum_{j=1, j\neq i}^{n} \delta_{ij}\left(\lambda_j^{\lambda_j^-/\lambda_j}\right)^2 I_j/\langle k\rangle^2 \quad (5)$$

### 3.2. Case retrieval

#### 3.2.1. Inference engine of decision case

Many key factors may take effect on management result while their complexity brings larger computation, so distinguishing special nodes from normal nodes is of great significance for reducing computational redundancy. These nodes reflect the special conditions and are defined as condition slots and restriction slots in complex network. The key factors have direct effect on decision result, so redundant similarity computation can be avoided by screening a large number of redundant cases. Among slots, restriction slots limit management condition. When restriction slots are within confine, some management measures are excluded. As condition slots can directly select the decision method, they avoid the redundant computation of similarity.

The key factors of restriction slots are algae area, investment of water treatment, long-term governance, second pollution and ecological security. These conclusions come from existing research and expert advice. The results for restriction slots of five key factors are translated into inference language according to inference rules as follows.

- If algae area >30 km$^2$, then D21, D22, D23 are excluded.
- If investment of water treatment <500 RMB/km$^2$, then D21, D22, D23, D41 are excluded.
- If long-term governance is required, then D21, D22, D23 are excluded.
- If second pollution is denied, then D33, D34 are excluded.
- If ecological security is required, then D11, D12 are excluded.

The key factors of condition slots are TP, TN, N/P, DO, pH and response time. The selection results for condition slots of six key factors are translated into inference language according to inference rules as follows:

- If TP > 2.00 mg/L OR TN > 2.00 mg/L, or 32:1 < N/P < 64:1, then D11, D31, D32, D42 are selected.
- If DO < 5 mg/L, then D11, D31, D32, D42 are selected. then D21 and D22 are selected.
- If 7.9 < pH < 8.1, then D33 and D34 are selected.
- If response time is required, then D21, D22 and D23 are selected.

In the inference rules, the meaning of $D_{ij}$ is shown in Fig. 4. This process can preliminarily screen the cases so that the

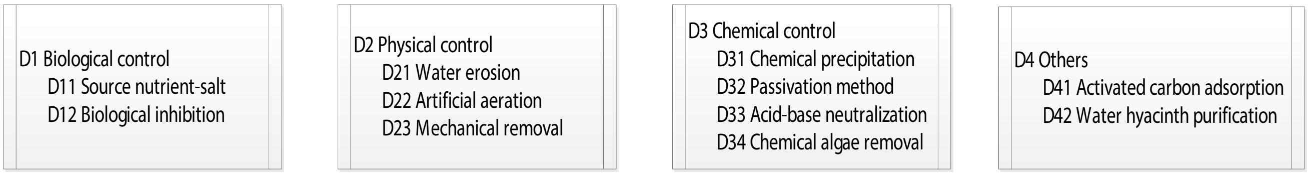


Fig. 4. Algae management methods.

computation of similarity is reduced. The design improves the efficiency of decision system and saves the response time.

### 3.2.2. Case matcher of decision-making

Nodes in complex network are classified into numerical, optional and Boolean types. For optional and Boolean nodes, the similarity is one when values of two cases are the same. Otherwise it is zero. For numerical nodes, they are hard to be identical because of uncertainty and fuzziness of environment. It may lead to the inaccurate similarity calculation. As a powerful tool for measuring similarity, IFRS expands and develops the fuzzy set. For algae management cases, it can better reflect the similarity among cases and explore the fuzzy concepts of factor value. We give an introduction of IFRS as follows. Let a set $S$ be fixed. If $\forall x \in S$, $X^+$ and $X^-$ denote upper approximation and lower approximation, respectively. An intuitionistic fuzzy set $A$ in $S$ is an object which has the form in Eq. (6) [28,29].

$$A = \left\{ \left\langle x, u_{A^-}(x), u_{A^+}(x), \gamma_{A^-}(x), \gamma_{A^+}(x) \right\rangle \middle| \forall x \in S \right\} \tag{6}$$

where $u_{A^-}: A^- \to [0,1], u_{A^+}: A^+ \to [0,1], \gamma_{A^-}: A^- \to [0,1], \gamma_{A^+}: A^+ \to [0,1]$. $u_{A^-}$ is called the membership degree function of lower approximation of the element $x \in S$ to $A$. It denotes certainty events of the negative impact from nodes. Its negativity decides the preference of decision process. $u_{A^+}$ is called the membership degree function of upper approximation. It denotes possibility events of the negative impact from nodes. Its negativity possibly decides the preference of decision process. $\gamma_{A^-}$ is called the non-membership degree function of lower approximation. It denotes certainty events of the positive impact from nodes. Its positivity decides the preference of decision process. $\gamma_{A^+}$ is called the non-membership degree function of upper approximation. It denotes possibility events of the positive impact from nodes. Its positivity possibly decides the preference of decision process. These variables all meet $0 \le \mu(x) + \gamma(x) \le 1$. Based on the function above, hesitation function $\pi_A(x)$ is defined to reflect the hesitation between membership and non-membership. It is calculated by $\pi_A(x) = 1 - u_A(x) - \gamma_A(x), \forall x \in S$. We propose normal distribution to denote membership functions. For example, membership functions of TP are as follows:

$$\gamma(\text{TP}) = \begin{cases} 0.9, \text{TP}_i \le 25 \\ 0.9\exp\left(-\left((x-25)/25\right)^2\right), \text{TP}_i > 25 \end{cases}$$
$$\mu(\text{TP}) = \begin{cases} 0.9\exp\left(-\left((x-200)/30\right)^2\right), \text{TP}_i < 200 \\ 0.9, \text{TP}_i \ge 200 \end{cases} \tag{7}$$

where $\gamma_{\text{TP}}$ and $\mu_{\text{TP}}$ are both lower approximation of IFRS while other nodes are in the same way. In IFRS of set $A$, intuitive index $\pi_A(x)$ is defined as the hesitation measure of $x$ to $A$ where $\pi_A(x) = 1 - u_A(x) - \gamma_A(x)$. For key factors in algae management, the direct calculation of similarity between numerical nodes is smaller than the reality due to the large uncertainty, so we modify the formula $\pi_A(x) = \alpha_A(1 - u_A(x) - \gamma_A(x))$ by introducing exponential operator $\alpha_A$.
For IFRS of set $A$ and $B$ in non-empty domain $X = \{x_1, x_2, \cdots, x_n\}$, the IFRS of set $A$ is $x_A = \left\langle u_{A^-}(x), u_{A^+}(x), \gamma_{A^-}(x), \gamma_{A^+}(x), \pi_A(x) \right\rangle$. The set $B$ is $x_B = \left\langle u_{B^-}(x), u_{B^+}(x), \gamma_{B^-}(x), \gamma_{B^+}(x), \pi_B(x) \right\rangle$. The similarity of $A$ and $B$ is defined as Eq. (8).

$$\begin{aligned} M(A,B) &= M(x_A, x_B) \\ &= 1 - \frac{1}{5}\Big( \omega_1(x_A, x_B)\left|u_{A^-}(x) - u_{B^-}(x)\right| + \omega_2(x_A, x_B)\left|u_{A^+}(x) - u_{B^+}(x)\right| \\ &\quad + \omega_3(x_A, x_B)\left|\gamma_{A^-}(x) - \gamma_{B^-}(x)\right| + \omega_4(x_A, x_B)\left|\gamma_{A^+}(x) - \gamma_{B^+}(x)\right| \\ &\quad + \omega_5(x_A, x_B)\left|\pi_{A^-}(x) - \pi_{B^-}(x)\right|\Big) \end{aligned} \tag{8}$$

where the weight $\omega_1(x_A, x_B)$, $\omega_2(x_A, x_B)$, $\omega_3(x_A, x_B)$, $\omega_4(x_A, x_B)$, $\omega_5(x_A, x_B)$ is added to improve the compliance. It weakens the effect of uncertainty and has the same effect with $\alpha_A$.

### 3.2.3. Inference engine of decision case

In practice, inaccessible values of key factors are common problem. If this situation happens, the decision system and algorithm does not work. To solve the problem, we propose the definition of comprehensive contribution. If the value of key factor is accessible, the contribution value is the similarity calculated by Eq. (8). If the value of key factor cannot be accessible, the contribution value is zero which means this key factor does not contribute to the matching result. The definition of comprehensive contribution is shown in Eq. (9).

$$S_{pq} = \sum_{i=1}^{m} w_i \times \left( \sum_{j=1}^{n} \left( w_{ij} \times M_{ij} \right) \right) \tag{9}$$

where $S_{pq}$ is the similarity of case $p$ and $q$. $w_i$ is the $i$th node weight in total network. $w_{ij}$ is the $j$th node weight in $i$th sub-network. $w_i$ and $w_{ij}$ is obtained by complex network of dynamic relation in section 3.1. $M_{ij}$ is the $j$th node contribution in $i$th sub-network. For all matching cases, the case with largest contribution is the best matching case.

## 3.3. Case selection, revision and reusing

To avoid the management failure of selected strategies, matching threshold $T$ is set. if $S_{pq} \ge T$, it means the matching case is available. The decision result can be applied in project after small adjustment. If $S_{pq}$ does not reach the threshold, no matching case is applicable. In this condition, the management finds the help from domain expert. This judgement guarantees the effectiveness of decision result. Finally, reusing cases are all added into cases library to provide reference for other decision.

## 4. Experiment

### 4.1. Study area

To verify the method, we exemplify Kunming Lake as the experimental lake. It is a landscape lake located in Summer Palace of Western Beijing, China. As the 39% percentage of the water area in Beijing city, it is the largest lake of Beijing, China. The maximum capacity of Kunming Lake is 4.5864 million km$^3$ and it accounts for 44.4% of total lake capacity of Beijing. The Kunming Lake has existed for over 3,500 years. For biological species, it is one of the richest lakes in Beijing. Kunming Lake and Longevity Hill in Summer Palace both derive from natural landscape of Xishan Mountain.

For the recent decades, algae bloom frequently outbreaks in Kunming Lake. We select the ecological event happened in August, 2008 as the typical experimental lake. The map of the study site is shown in Fig. 5. The location of data acquisition is next to the Shi Qi Kong Qiao as the red dot shows. The related information of Kunming Lake is described in Table 1 in which $V_{ij}$ is the key factors in Fig. 3.

We also build the case library of algae bloom in Table 2. The case library includes some cases of the recent decades which provide the reference for algae management in the future. The case library is used in case matcher process and some information of case library is shown in Table 2.

### 4.2. Result and discussion

The key factors in Table 1 correspond with nodes in complex network. By calculating the weights of nodes in complex network, we can obtain the weights of key factors in Table 1. The network structure has been shown in Fig. 3, so the statistical characteristic parameters of complex network are as follows:

$$\lambda = \begin{bmatrix} 2 & 3 & 1 & 2 & 2 \end{bmatrix};\ \lambda^+ = \begin{bmatrix} 2 & 2 & 0 & 1 & 0 \end{bmatrix};\ \lambda^- = \begin{bmatrix} 2 & 1 & 1 & 1 & 2 \end{bmatrix} \tag{10}$$

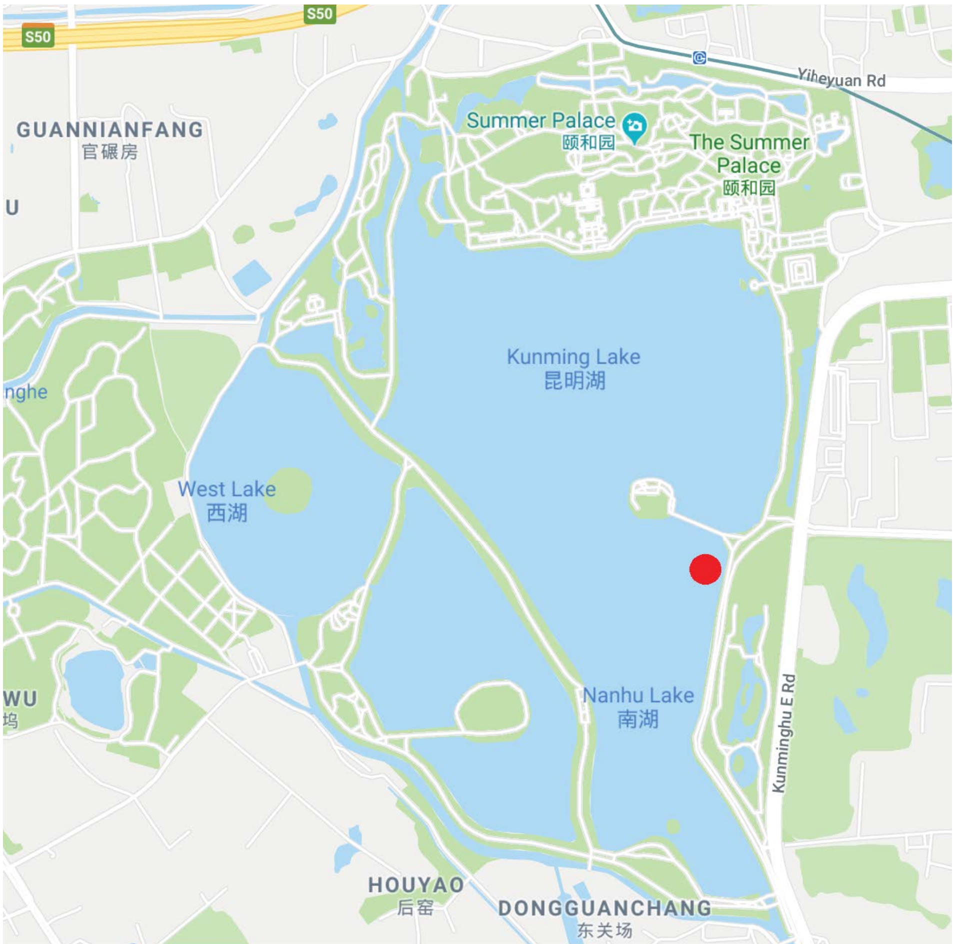


Fig. 5. Map of the study site.

Table 1
Information of Kunming Lake

| Name | Time | V1 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | V11 | V12 | V13 | V14 (km²) | V15 | V16 | – | – |
| Kunming Lake | August, 2008 | Light green | Cyanobacteria | Moderate | – | Viscous | Smelly | – | – |
| V2 | | | | | | | | | |
| V21 (mg/L) | V22 (mg/L) | V23 | V24 | V25 (S/m) | V26 (°C) | V27 (mg/L) | V28 (mg/L) | V29 (mg/L) | V210 (mg/L) |
| 63 | 2.05 | 32.54 | – | – | – | 8.9 | 2.5 | – | 30 |
| | | V3 | | | | V4 | | | |
| V31 | V32 | V33 | V34 | V35 | V41 | V42 | V43 | V44 | V45 |
| Quick | – | No | No | No | Medium | Low | High | Low | Low |
| V5 | | | | | | | | | |
| V51 (km²) | V52 | V53 | V54 | V55 (°C) | V56 | V57 | | | V58 |
| 1.94 | Landscape lake | Plain | – | – | – | N39°59′36.46″ E116°16′6.26″ | | | – |

Table 2
Information of case library

| No. | Name | Time | Area (km²) | Eutrophication level |
|---|---|---|---|---|
| 1 | Tai Lake | 2015.8.6 | 183 | Moderate |
| 2 | Tai Lake | 2016.5.27 | 412 | Moderate |
| 3 | Chao Lake | 2010.7.22 | 50 | Moderate |
| 4 | Dian Lake | 2008.7.30 | 181.09 | Severe |
| 5 | Tai Lake | 2018.6.17 | 80 | Moderate |
| 6 | Dongting Lake | 2008.7 | 10 | Moderate |
| 7 | Dianshan Lake | 2008.8.26 | 15 | Moderate |
| 8 | West Lake | 2010.7 | – | Mild |
| 9 | Xuanwu Lake | 2005.9.18 | 3 | Moderate |
| 10 | Dong Lake | 2009.8.16 | 4 | Moderate |
| 11 | Nanji Wetland | 2010.10.19 | 5.12 | Moderate |
| 12 | South water reservoir | 2009.2.6 | – | Moderate |

$$d=\begin{bmatrix}0&1&3&2&1\\1&0&2&1&1\\3&2&0&1&3\\2&1&1&0&2\\1&1&3&2&0\end{bmatrix}\qquad \delta=\begin{bmatrix}1&1&0&0&1\\1&1&0&1&1\\0&0&1&1&0\\0&1&1&1&0\\1&1&0&0&1\end{bmatrix} \tag{11}$$

These parameters are processed by Eq. (2). The node efficiency and weight are obtained as follows:

$$I=\begin{bmatrix}\frac{7}{5}&1&\frac{9}{5}&\frac{6}{5}&\frac{7}{5}\end{bmatrix} \tag{12}$$

For the total network, the node weights are $W_V=\begin{bmatrix}0.2439&0.3085&0.0980&0.1056&0.2439\end{bmatrix}$ according to Eq. (5).

For the sub-network, the node weights are as follows according to Eq. (5).

$$\begin{aligned}
W_{V1}&=\begin{bmatrix}0.1798&0.0877&0.2468&0.1212&0.1204&0.2441\end{bmatrix}\\
W_{V2}&=\begin{bmatrix}0.0499&0.0499&0.2222&0.2282&0.0426&0.0614&0.1072&0.0119&0.0119&0.2148\end{bmatrix}\\
W_{V3}&=\begin{bmatrix}0.2122&0.1143&0.2898&0.2640&0.1196\end{bmatrix}\\
W_{V4}&=\begin{bmatrix}0.0524&0.2590&0.0679&0.3104&0.3104\end{bmatrix}\\
W_{V5}&=\begin{bmatrix}0.0162&0.1648&0.0622&0.2060&0.2491&0.1218&0.0372&0.1427\end{bmatrix}
\end{aligned} \tag{13}$$

The node weights above further denote the weight of key factors in Table 1. On the basis, the values of key factors are used to calculate the similarity value. When the case presentation is ready, the decision process comes into inference engine as described in section 3.2.1.

For restriction slots, only second pollution is implemented. So acid–base neutralization and chemical algae removal are excluded. For condition slots, TN, N/P and response time are implemented. So, source nutrient-salt, water erosion, artificial aeration and mechanical removal are selected. These methods can reduce large amount of computation with less cases in the next step. Besides, the key factors which have been implemented will not work then. Therefore five cases are selected from case library and they are case 2 (basin A in Tai Lake in May, 2016, N31°30′31.16″, E120°11′14.48″), case 3 (Chao Lake in 2010, N31°36′50.00″, E117°47′28.20″), case 5 (basin B in Tai Lake in June, 2018, N30°56′33.39″, E120°10′4.30″), case 8 (West Lake in July, 2007, N30°15′8.27″, E120°09′13.48″), case 11 (Nanji Wetland in October, 2010, N28°57′17.42″, E116°20′28.26″). These cases have similar situation of management requirements and water environment, so the management method of Kunming Lake can imitate one of these cases. To select the most analogous case, we conduct the experiment process according to section 3.2.2 and calculate the contribution of each case, respectively. The comprehensive values are from the similarity of each key factor, so we first calculate the similarity of each factor. We exemplify the key factor TP to calculate the similarity. The TP of five cases are $TP_2 = 210$, $TP_3 = 150$, $TP_5 = 130$, $TP_8 = 147$, $TP_{11} = 250$ and the TP value of Kunming Lake is $TP_D = 63$. Correspondingly, the IFRS values $x_A = \langle u_{A^-}(x), u_{A^+}(x), \gamma_{A^-}(x), \gamma_{A^+}(x), \pi_A(x) \rangle$ are as follows:

$$\begin{aligned}
&TP_D = \langle 0 \quad 0.9107 \quad 0.0893 \quad 1 \quad 0.9107 \rangle \\
&TP_2 = \langle 0.9 \quad 1 \quad 0 \quad 0.1 \quad 0.1 \rangle \\
&TP_3 = \langle 0.056 \quad 1 \quad 0 \quad 0.944 \quad 0.944 \rangle \\
&TP_5 = \langle 0.0039 \quad 1 \quad 0 \quad 0.9961 \quad 0.9961 \rangle \\
&TP_8 = \langle 0.0397 \quad 1 \quad 0 \quad 0.9603 \quad 0.9603 \rangle \\
&TP_{11} = \langle 0.9 \quad 1 \quad 0 \quad 0.1 \quad 0.1 \rangle
\end{aligned} \tag{14}$$

We assume that $\omega_1(x_A,x_B) = 1.5$, $\omega_2(x_A,x_B) = 0.5$, $\omega_3(x_A,x_B) = 1.5$, $\omega_4(x_A,x_B) = 0.5$, $\omega_5(x_A,x_B) = 1$ and $\alpha_A = 0.5$. Through Eq. (8), the values of similarity are follows:

$$\begin{aligned}
&M(TP_D, TP_2) = 0.5232;\ M(TP_D, TP_3) = 0.9386; \\
&\qquad M(TP_D, TP_5) = 0.9542 \\
&M(TP_D, TP_8) = 0.9434;\ M(TP_D, TP_{11}) = 0.5232
\end{aligned} \tag{15}$$

The values of comprehensive contribution are shown as follows and they can be depicted in Fig. 6.

$$S_{D2} = 0.3235,\quad S_{D3} = 0.4794,\quad S_{D5} = 0.4358,\quad S_{D8} = 0.5231,\quad S_{D11} = 0.3761 \tag{16}$$

Case 8 has the largest matching value from the result in Fig. 6, so it is the best matching case. By learning the decision method in case library, the method of case 8 is mechanical removal. This method pumps the algae-water into filter tank by electromechanical equipment. The equipment can separate the algae from water by filtering. This method improves the water purification in a short time and lowers the algae content which prevents algae growth effectively. It provides a good reference for Kunming Lake. If it is strictly adopted and implemented in reality, the management process and effect should be documented and retained in case library. The retaining can guarantee the integrity, coverage and richness of case library.

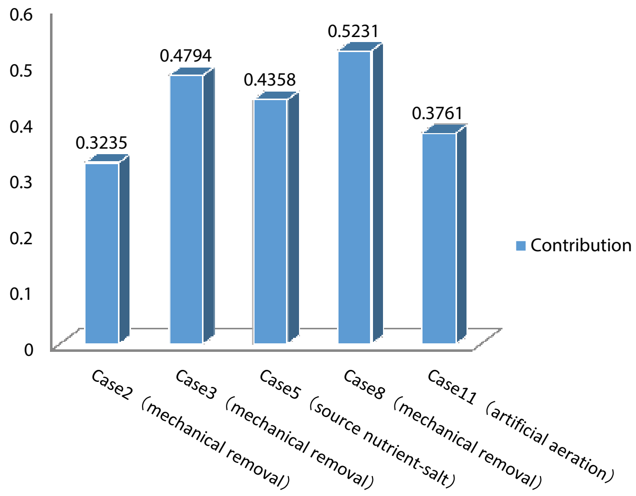


Fig. 6. Contribution comparison of matching cases.

In the case study, the case of Kunming Lake is the case which did happen. According to document, the decision association took measure of mechanical removal when the algae bloom out broke in 2008. It accords with the reasoning result. In addition, we analyze the outbreak condition and management requirement issued by environmental protection department and conclude that the algae outbreak of Kunming Lake is most similar to West Lake. Therefore the method in our paper can accurately retrieve the measures which accord with preset result.

To further reflect the retrieval result, we rank the strategies of algae management based on CBR-OSS model and compare the trend with Vague set multi-objective decision making, Fuzzy Bayes and Text analysis in Fig. 7. The figure shows that mechanical removal is a most suitable strategy for Kunming Lake by four methods. It means mechanical removal is the first choice while artificial aeration and source nutrient-salt are following. The rank trend of four methods is consistent of the result in Fig. 6.

Accuracy, operation time, expert participation and flexibility are four factors to evaluate the method advantage of algae management decision. Accuracy denotes the proportion of right decisions. The operation time is from the simulation in MATLAB. Expert participation is the proportion of expert decision in the whole decision process. Flexibility denotes the completion of the decision process when some information is missed. If accuracy and flexibility are higher, the decision model is better. If operation time and expert participation are lower, the decision model is better. We select another 20 lakes and reservoirs worldwide

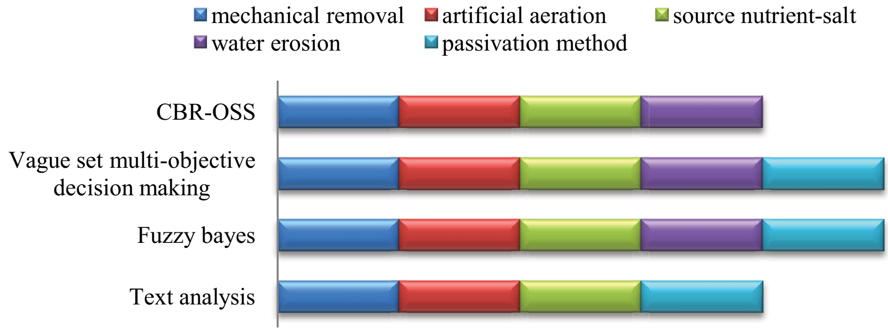


Fig. 7. Rank trend of four methods.

to verify the CBR-OSS model which are simulated in the MATLAB. These performance indexes are shown in Table 3 and Fig. 8.

Fig. 8 shows that CBR-OSS is more accurate and flexible than Text analysis, fuzzy Bayes and vague set multi-objective. Although this method is not better than expert opinion in accuracy and flexibility, it can save time and replace experts to some extent.

Considering the complexity of water environment and incomplete decision in algae management, we propose CBR-OSS model. In the experiment, the algae bloom in Kunming Lake did happen and the decision was from the experts. The selected strategy has been applied. By CBR-OSS model, the same decision can be obtained. The rank trend is consistent with other three methods. To further analyze the characteristics of CBR-OSS model, we compare it with expert decision, Text analysis, Fuzzy Bayes and vague set multi-objective in the aspect of accuracy, operation time, expert participation and flexibility. The comparison result shows that:

- The decision result by CBR-OSS model is credible. Its rank trend and accuracy suggest the model has combined all the key information and gives the reasonable decision. The reasoning process is based on real case with their decision result documented in case library. Its possible outcome can be expected, so emergency and decision consequences are predictable.
- CBR-OSS model reduces the expert participation by imitating the thinking way of human beings. It has abstracted the complex factors and forms the dynamic complex network to reflect the intrinsic interaction.

Table 3
Comparison of performance parameters

| Performance indexes | Expert decision | Text analysis | Fuzzy Bayes | Vague set multi-objective decision making | CBR-OSS |
|---|---|---|---|---|---|
| Number | (1) | (2) | (3) | (4) | (5) |
| Accuracy | 100% | 85% | 70% | 80% | 90% |
| Operation time (s) | – | 5 | 1.6 | 2.9 | 2.8 |
| Expert participation | 100% | 70% | 0 | 40% | 20% |
| Flexibility | 100% | 80% | 0 | 70% | 90% |

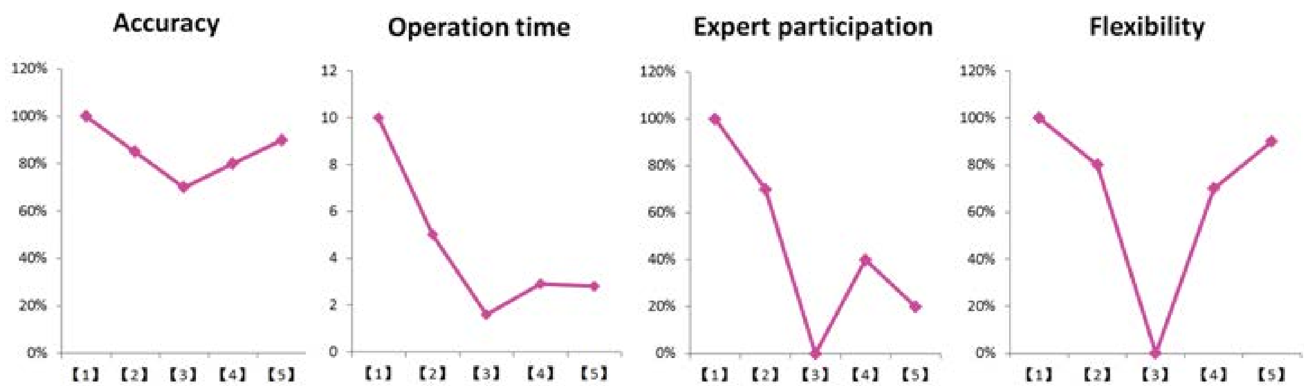


Fig. 8. Comparison of performance parameters.

The resulting uncertainty and randomness are finally processed by IFRS.
- The obstacle of inaccessible data can be solved by the definition of contribution which makes the CBR-OSS model practical and flexible. The inference engine can largely reduce the operation time which is shown in the experiment.
- The decision method has strong operability. It can further refer to operation procedures, reagent and dosage from matching case.

## 5. Conclusion

The paper lists the problems of algae management and the limitation of intelligent decision system. To solve the problem, the CBR-OSS model is proposed. This model introduces thinking pattern of CBR to determine the method of algae management. The model also combines dynamic complex network and IFRS to reflect the complex mechanism of algae outbreak in the decision process. As the experiments show, CBR-OSS model outperforms in accuracy, operation time, expert participation and flexibility. This technique strengthens the utilization of intelligent information and realizes the automatic data processing. It promotes the information processing and automation technology in environment protection.

## Acknowledgment

This work has been supported by the National Natural Science Foundation of China (No. 61573061, 61703008 and 61802010).

## Symbols

CBR-OSS — Case-based reasoning-optimal strategy selection
IFRS — Intuitionistic Fuzzy rough sets
HABs — Harmful algae blooms
OSS — Optimal strategy selection
CBR — Case-based reasoning
MADM — Multiple attribute decision making
AI — Artificial intelligence
TP — Total phosphorus
TN — Total nitrogen
N/P — Ratio of nitrogen and phosphorus
EC — Electronic conductivity
TW — Temperature of water
DO — Dissolved oxygen
BOD — Biochemical oxygen demand
COD — Chemical oxygen demand

## References

[1] W. Rast, J.A. Thornton, Trends in eutrophication research and control, Hydrol. Processes, 10 (2015) 295–313.
[2] Y. Chen, S. Yu, S. Tang, Y. Li, H. Liu, X. Zhang, G. Su, B. Li, H. Yu, J.P. Giesy, Site-specific water quality criteria for aquatic ecosystems: a case study of pentachlorophenol for Tai Lake, China, Sci. Total Environ., 541 (2016) 65–73.
[3] M.W. Matthews, Eutrophication and cyanobacterial blooms in South African inland waters: 10 years of MERIS observations, Remote Sens. Environ., 155 (2014) 161–177.
[4] R. Loos, B.M. Gawlik, G. Locoro, E. Rimaviciute, S. Contini, G. Bidoglio, EU-wide survey of polar organic persistent pollutants in European river waters, Environ. Pollut., 157 (2008) 561–568.
[5] J.E. Sawyers, A.S. McNaught, D.K. King, Recent and historic eutrophication of an island lake in northern Lake Michigan, USA, J. Paleolimnol., 55 (2016) 97–112.
[6] G.T. Frumin, I.M. Gildeeva, Eutrophication of water bodies — a global environmental problem, Russ. J. Gen. Chem., 84 (2015) 2483–2488.
[7] P. Jia, Y. Zhou, X. Zhang, Y. Zhang, R. Dai, Cyanobacterium removal and control of algal organic matter (AOM) release by UV/$H_2O_2$ pre-oxidation enhanced Fe(II) coagulation, Water Res., 131 (2018) 122–130.
[8] M. Huang, B. Li, C. Zhang, Z. Zhang, Z. Lei, B. Lu, B. Zhou, Effect of algae growth on aerobic granulation and nutrients removal from synthetic wastewater by using sequencing batch reactors, Bioresour. Technol., 179 (2015) 179–192.
[9] M.R. Sengco, Prevention and control of *Karenia brevis* blooms, Harmful Algae, 8 (2009) 623–628.
[10] H. Liang, J. Nan, W.-j. He, G. Li, Algae removal by ultrasonic irradiation–coagulation, Desalination, 239 (2009) 191–197.
[11] K. Lee, K.-S. Jang, S. Kim, S.-W. Park, Performance assessment of apparatus for controlling algae bloom in aqua pet pank using by a cold plasma, J. Korean Soc. Fish. Ocean Technol., 49 (2013) 461–465.
[12] D. Wei, Y. Tao, Z. Zhang, L. Liu, X. Zhang, Effect of *in-situ* ozonation on ceramic UF membrane fouling mitigation in algal-rich water treatment, J. Membr. Sci., 498 (2015) 116–124.
[13] K. Wang, J. Saththasivam, W. Yiming, K. Loganathan, Z. Liu, Fast and efficient separation of seawater algae using a low-fouling micro/nano-composite membrane, Desalination, 433 (2018) 108–112.
[14] H. Wei, J. Shang, L. Xin, C. Fan, Effects of sludge dredging on the prevention and control of algae-caused black bloom in Taihu Lake, China, Acad. J. Environ. Sci., 25 (2013) 430–440.
[15] S. Gao, M. Du, J. Tian, J. Yang, Y. Yang, F. Ma, J. Nan, Effects of chloride ions on electro-coagulation-flotation process with aluminum electrodes for algae removal, J. Hazard. Mater., 182 (2010) 827–834.
[16] W. Hu, C. Wu, A. Jia, Z. Zhang, F. Chen, Impact of diatomite on the slightly polluted algae-containing raw water treatment process using ozone oxidation coupled with polyaluminum chloride coagulation, Environ. Technol., 35 (2014) 2670–2675.
[17] F.M.S.E. El-Dars, M.A.M. Abdel Rahman, O.M.A. Salem, E.-S.A. Abdel-Aal, Algal control and enhanced removal in drinking waters in Cairo, Egypt, J. Water Health, 13 (2015) 1060–1072.
[18] Q. Shen, J. Zhu, L. Cheng, J. Zhang, Z. Zhang, X. Xu, Enhanced algae removal by drinking water treatment of chlorination coupled with coagulation, Desalination, 271 (2011) 236–240.
[19] X. Ji, M. Jiang, J. Zhang, X. Jiang, Z. Zheng, The interactions of algae-bacteria symbiotic system and its effects on nutrients removal from synthetic wastewater, Bioresour. Technol., 247 (2018) 44–50.
[20] S.P. Zuo, K. Wan, S.M. Ma, L.T. Ye, Combined allelopathic potential of aquatic plants species to control Algae, Allelopathy J., 34 (2014) 315–324.
[21] Z. Liang, Y. Liu, F. Ge, N. Liu, M. Wong, A pH-dependent enhancement effect of co-cultured *Bacillus licheniformis* on nutrient removal by *Chlorella vulgaris*, Ecol. Eng., 75 (2015) 258–263.
[22] Z. Liu, L. Li, X. Wang, C. Chen, Researches of water bloom emergency management decision-making method and system based on fuzzy multiple attribute decision making, Int. J. Comput. Sci., 9 (2012) 35–42.
[23] Y.T. Bai, L. Wang, X.Y. Wang, The research of decision-making method for multi-objective in water bloom emergency governance based on vague set theory, J. Comput. Info. Sys., 10 (2014) 2099–2106.
[24] Y.-T. Bai, B.-H. Zhang, X.-Y. Wang, X.-B. Jin, J.-P. Xu, T.-L. Su, Z.-Y. Wang, A novel group decision-making method based on sensor data and fuzzy information, Sensors, 16 (2016) 1799–1806.

[25] Y.-T. Bai, B.-H. Zhang, X.-Y. Wang, X.-B. Jin, J.-P. Xu, Z.-Y. Wang, Expert decision support technique for algae bloom governance in urban lakes based on text analysis, Water, 9 (2017) 308–315.
[26] J. Zhang, J. Shi, X. Chang, A mathematical model of algae growth in a pelagic-benthic coupled shallow aquatic ecosystem, J. Math. Biol., 76 (2017) 1–35.
[27] F. Shao, X.Y. Wang, Y. Shi, J. Xu, S. Lu, W. Li, T. Lina, Research on recognition method of algae blooms based on complex network, J. Appl. Sci., 13 (2013) 1847–1850.
[28] C. Cornelis, M. De Cock, E.E. Kerre, Intuitionistic fuzzy rough sets: at the crossroads of imperfect knowledge, Expert Syst., 20 (2010) 260–270.
[29] B. Sun, W. Ma, Q. Liu, An approach to decision making based on intuitionistic fuzzy rough sets over two universes, J. Oper. Res. Soc., 64 (2013) 1079–1089.