

Article

Expert Decision Support Technique for Algal Bloom Governance in Urban Lakes Based on Text Analysis

Yu-Ting Bai ¹, Bai-Hai Zhang ^{1,*}, Xiao-Yi Wang ², Xue-Bo Jin ², Ji-Ping Xu ² and Zhao-Yang Wang ¹

¹ School of Automation, Beijing Institute of Technology, Beijing 100081, China; byting@bit.edu.cn (Y.-T.B.); wangzhaoyang@bit.edu.cn (Z.-Y.W.)

² School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China; wangxy@btbu.edu.cn (X.-Y.W.); jinxuebo@btbu.edu.cn (X.-B.J.); xujp@th.btbu.edu.cn (J.-P.X.)

* Correspondence: smczhang@bit.edu.cn; Tel.: +86-139-1068-9145

Academic Editor: Peter Coombes

Received: 15 February 2017; Accepted: 24 April 2017; Published: 28 April 2017

Abstract: As a typical phenomenon of eutrophication pollution, algal bloom threatens public health and water security. The governance of algal bloom is largely affected by administrators' knowledge and experience, which may lead to a subjective and one-sided decision-making result. Meanwhile, experts in the specific field can provide professional support. How to utilize expert resources adequately and automatically has been a problem. This paper proposes an expert decision support technique for algal bloom governance based on text analysis methods. Firstly, the decision support mechanism is introduced to form a general decision-making framework. Secondly, the expert classification method is proposed to help with choosing suitable experts. Thirdly, a multi-criteria group decision-making method is presented based on the automatic analysis of experts' decision opinions. Finally, an experiment is conducted to verify the expert decision support technique. The results show the technique's feasibility and rationality. This paper describes experts' information and opinions with natural language, which can intuitively reflect the natural meaning. The expert decision support technique based on text analysis broadens the management thought of water pollution in urban lakes.

Keywords: algal bloom; eutrophication; urban lake; governance decision-making; text analysis

1. Introduction

Water pollution has been caused by human beings' complicated activities. In urban lakes, eutrophication is a main pollution created by excessive nutrients such as nitrogen and phosphorus [1,2]. Nutrients lead to the rapid breeding of algae and phytoplankton, which consume vast dissolved oxygen. As a result, water quality degrades and fish die, as do other aquatic creatures. Globally, nearly 40% of lakes and rivers suffer from different levels of eutrophication. In China, 22.9% of the major lakes and reservoirs are still eutrophicated; among these, Lake Dianchi has been under medium eutrophication for several years [3]. In Beijing, 86.4% of lakes are under mild or medium eutrophication [4]. Algal bloom is a typical phenomenon of eutrophication [5,6]. It broke out on a large scale in many major lakes, such as Lake Taihu in Jiangsu Province, Lake Dianchi in Yunnan Province, Lake Chaohu in Anhui Province, and Lake Beihai in Beijing. Although many governance measures have been adopted, algal bloom still breaks out in urban lakes, for example, River Chuhe in Wuhan City was covered by algal bloom in a 2 km scale in August 2016 [7]. Moreover, researchers analyzed and proved the algal bloom hazards with experimental data and cases [8,9]. As a conclusion, algal bloom destroys urban landscapes and is a serious threat to public health.

For the serious algal bloom hazards, various governance approaches have been studied and applied, including physical, chemical, biological and ecological methods [10–12]. The related research focuses on the microcosmic principles, such as destroying algae's cysts, removing flottage and changing nutritive element content. However, their applicable condition and selection basis have not been studied specifically. Moreover, many administrators select governance approaches only depending on personal knowledge and experience. Some decision-making methods have been proposed to solve the selection problem [13–17]. Their basic thought is to synthesize the evaluation information from different aspects. In the evaluation, decision experts give opinions in the form of grade scores or variables, which may misrepresent experts' original thought. Besides, studies barely focus on the selection of decision experts, and the importance weights of experts are not given in a rational computing method.

In the ideal management of water pollution, governance approaches should be selected based on their adaptation to the current environment and management objectives. The adaptation information should be supported adequately by experts in the specific field. And the information should be analyzed efficiently with advanced computer and big data techniques. In this context, we propose an expert decision support mechanism and methods based on text analysis for algal bloom governance. First, experts studying algal bloom are classified by analyzing their profiles and literature lists with the fusion algorithm, which combines keyword extraction with classification neural network. Classification results reflect experts' professional level and help in selecting the final decision experts. Second, the selected experts are invited to provide decision opinions in the form of natural language texts. Third, decision opinions are processed to form decision matrix, and the final result is given with a multi-criteria group decision-making method. In this paper, we provide the selecting basis of experts by automatically analyzing their information. Decision opinions are expressed with natural language to reflect experts' original thoughts in a high extent. The decision-making experiment on algal bloom governance indicates that the new decision mechanism is feasible, and the text analysis broadens the thought of water pollution management.

This paper is organized as follows: Section 2 introduces the general decision support mechanism firstly. Then the expert classification method is introduced based on forward maximum matching word segmentation algorithm (FMM) and self-organizing-map neural network (SOM). The multi-criteria group decision-making method is introduced based on the processing of decision opinions in natural language. Section 3 conducts the decision-making experiment on algal bloom governance in urban lakes. Section 4 discusses the main work. Section 5 gives the conclusion of this paper.

2. Expert Decision Support Technique

2.1. Decision Support Mechanism

To modify the existing decision-making mechanism, which strongly depends on personal knowledge and experience, an expert decision support mechanism is proposed for algal bloom governance in urban lakes. The mechanism consists of expert database formation, decision opinion processing, final decision-making and basic information database update, as shown in Figure 1.

The deficiency of specialized knowledge limits administrators' efficient and refined management. Meanwhile, experts in the specialized field can be the significant supplemental resources. As the basis of decision support mechanism, the expert database gathers experts with their information analyzed automatically. Their profile texts and literature lists are processed with FMM–SOM method to classify the experts.

When algal bloom needs precaution or governance, some experts in the database are invited to provide decision opinions. The opinions are the important reference for next steps and are in the form of natural language to reflect natural thoughts in a great extent. The natural language is transformed with text analysis methods to be used in the decision-making calculation.

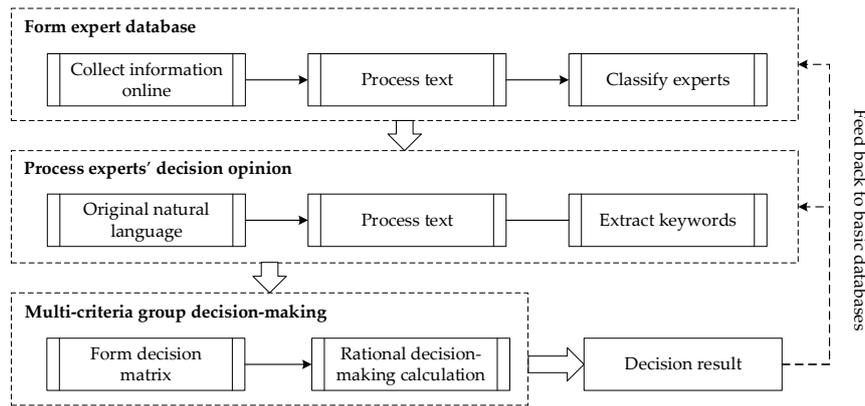


Figure 1. The flow and constitution of expert decision support mechanism for algal bloom governance.

The processed experts’ opinions are applied to multi-criteria group decision-making, which can pick out suitable approaches. Administrators can make the final decision considering the results. Each cycling of the decision mechanism can produce new information, which can update the basic databases, such as professional dictionaries and keyword sets for text analysis.

In the expert decision support mechanism, expert classification and opinion analysis are all processed with calculable methods. Therefore, it can be realized automatically in the computer system with new information technology. The mechanism can increase the efficiency and professionalism of algal bloom governance.

2.2. Expert Classification Based on FMM–SOM

For computer comprehension of natural language, the first step is to segment sentences into words. FMM is selected to process the Chinese sentences, which can select the longest compound words from the dictionary [18]. It can read in and segment the Chinese character strings circularly. The algorithm flow is shown in Figure 2.

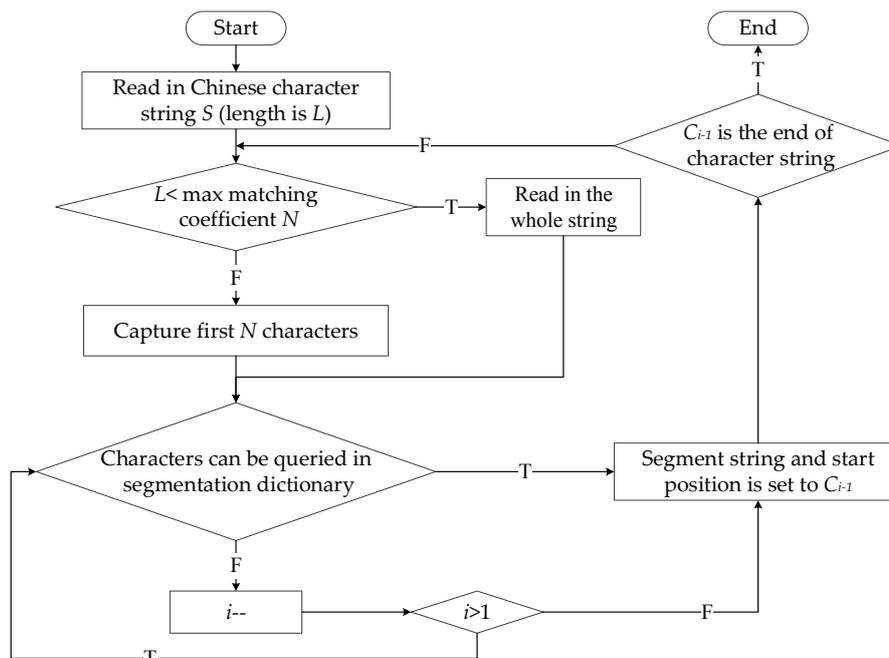


Figure 2. The flow of forward maximum matching word segmentation algorithm (FMM). The initial $i = L$, and C_i is the series number of the character.

SOM is chosen to classify objects automatically. The classical SOM can obtain the classification information of input sources [19], while it does not produce feedback error between the classified and anticipated results. The assessment of classification results mainly depends on the subjective judgement and lacks the quantitative support. Therefore, a matching degree is applied into SOM.

Assume that there are n elements in the input vector and k output nodes, which means that n objects need to be classified to k classes. The anticipated and actual classification results are expressed as k sets G_i and T_j , respectively, in which $G_i = \{g_{i,1}, g_{i,2}, \dots, g_{i,x(i)}\}$, $T_j = \{t_{j,1}, t_{j,2}, \dots, t_{j,y(j)}\}$, $i = 1, 2, \dots, k, j = 1, 2, \dots, k, \sum_{i=1}^k x(i) = n, \sum_{j=1}^k y(j) = n$. The matching degree S is defined as:

$$S = \begin{cases} MAX & \text{if } num(P(G_i, T_j) = 1) = k \\ \sum_{i=1}^k \sum_{j=1}^k P(G_i, T_j) & \text{if } num(P(G_i, T_j) = 1) < k \end{cases} \quad (1)$$

where MAX means the complete matching, otherwise, the higher the degree S , the better the classification result. $num()$ is a counting function. $P(G_i, T_j)$ means the overlap ratio of the anticipated and actual results, $P(G_i, T_j) = num(Rep(G_i, T_j)) / x(i)$ in which $Rep()$ is used to find out the overlapped elements between G_i and T_j , and $x(i)$ is the number of elements in G_i .

The matching degree S indicates the similarity between the anticipated and actual classification results. It can be obtained in each running of SOM with different training times. The best network structure and training time can be confirmed by judging the matching degree S .

Based on the preparation of FMM and SOM, the fusion algorithm is proposed to classify the experts based on text inputs. The concrete procedure of the algorithm is as follows:

- (1) The to-be-classified object set is $V = \{v_1, v_2, \dots, v_n\}$, and each object v_i is described with a text of $s(i)$ Chinese sentences. The texts are denoted as word set $Y_i = \{y_{i,1}, y_{i,2}, \dots, y_{i,s(i)}\}$. A keyword set $KW = \{kw_1, kw_2, \dots, kw_p\}$ is introduced to filter the texts.
- (2) Each sentence $y_{i,s(i)}$ is segmented with FMM, and the word segmentation results are denoted as $DW_i = \{dw_{i,1}, dw_{i,2}, \dots, dw_{i,w(s)}\}$, $s = 1, 2, \dots, s(i)$, and $w(s)$ is the number of segmentation words of the s -th sentence.
- (3) For the word segmentation result DW_i , all of its elements are scanned to match the keywords in the set KW . The repeated times of each keyword in DW_i are recorded in the attribute matrix of the to-be-classified object:

$$M_{att} = \begin{matrix} & & kw_1 & \dots & kw_p \\ \begin{matrix} v_1 \\ \vdots \\ v_n \end{matrix} & \begin{bmatrix} r_{11} & \dots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{n1} & \dots & r_{np} \end{bmatrix} \end{matrix} \quad (2)$$

where r_{ij} indicates the repeated times of the j -th keyword in the i -th word segmentation result DW_i of the object v_i , and $r_{ij} = 0$ if the j -th keyword does not appear in DW_i .

- (4) Set that there are p nodes in input layer of SOM network. The elements in each row of attribute matrix M_{att} are set as the input of SOM network. The input data are trained by different times and the matching degree S is calculated in each situation. The suitable network model is selected by judging the matching degree S . Then test objects can be classified with the selected SOM network.

2.3. Multi-Criteria Group Decision-Making Based on Natural Language

When or before algal bloom breaks out, the suitable governance approach should be selected. In the proposed expert decision support mechanism, experts are invited to provide decision opinions using natural language. The language information is processed as the input of the multi-criteria group decision-making method. The processing and decision-making methods are introduced in this section.

2.3.1. Keyword Extraction of Opinions

Experts' opinions are expressed in natural language which reflects personal thoughts more intuitively. However, its automatic processing and accurate comprehension are strongly complicated. The basic processing method is proposed to extract the key information from opinions.

The opinion processing includes the word segmentation, part-of-speech tagging, word filtering and semantic similarity computation. The FMM algorithm is used to segment words, and the other procedures are introduced as follows.

Part-of-Speech Tagging

The words are clustered into different classes according to the similar syntactic structure and classical semantic type. The classes are called the part of speech, which can provide useful information of the word and its adjacent components. The common part-of-speech includes the noun, verb, pronoun, article, etc. The part-of-speech tagging sets have different rules and standards for Chinese. We choose the mature corpus "The People's Daily' Tagged Corpus" [20] which has been widely used in automatic language understanding. The Viterbi algorithm [21] is used as the tagging method, which is universal in different language environments and can keep the algorithm efficiency when words increase.

Word Filtering

There are many conjunctions in Chinese sentences, like the words *and*, *or*, *for*, *of* in English. They appear frequently but have little meaning for the text understanding. The useless words are filtered using the stop word list to eliminate their interference in the keyword extraction. In the concrete procedure, each word in the segmentation results is judged as to whether it exists in the stop word list, and it is deleted if it exists.

Semantic Similarity Computation

Semantic similarity is the key factor of keyword extraction and is usually described with the coding distance between words in a synonym dictionary. "Chinese Thesaurus (Extended)" from Harbin Institute of Technology (HIT) [22] is set as the basic synonym dictionary in this paper. Each word in the dictionary has several codings, and each coding is described with five layer codes and a flag bit. The coding $Code_i = X_{i1}X_{i2}X_{i3}X_{i4}X_{i5}F_i$. The five layer codes represent the broad class, medium class, narrow class, word group and atomic word group. The flag bit can be "=", "#", or "@", in which "=" means synonymous, "#" means the same class and "@" means the word is independent and self-enclosed because it has no synonymous or related words.

The semantic distance of words W_1 and W_2 is defined as

$$Dis(W_1, W_2) = \min_{i=1,2,\dots,m; j=1,2,\dots,n} Dis(Code_{1i}, Code_{2j}) \quad (3)$$

where the word W_1 has m coding ($Code_{1i}$), and the word W_2 has n coding ($Code_{2j}$). $Dis()$ means the distance between the coding $Code_{1i}$ and $Code_{2j}$, and it is measured with the concept hierarchy tree:

$$Dis(Code_1, Code_2) = \begin{cases} 0 & Code_1 = Code_2 \text{ and } F_1 = F_2 \neq "@" \\ weights[5] \times init_dis & Code_1 = Code_2 \text{ and } F_1 = F_2 = "@" \\ weights[i-1] \times init_dis & Codes \text{ begin from } i-th \text{ layer} \end{cases} \quad (4)$$

where $weights[]$ means the semantic distance weights of different layers in the concept hierarchy tree. $weights = [\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_F]$, $\omega_1 > \omega_2 > \omega_3 > \omega_4 > \omega_5 > \omega_F$. The higher the layer, the larger the weight. $init_dis$ means the custom initial distance between two codes in the tree.

Then we give the semantic similarity between W_1 and W_2 :

$$Sim(W_1, W_2) = \frac{\alpha}{Dis(W_1, W_2) + \alpha} \tag{5}$$

where α means the word distance when the similarity is 0.5 and it can be adjusted. α decides the value range of similarity. The larger the α , the more insensitive the semantic similarity. $\alpha = 5$ and the value range of similarity is 0.33~1 in this paper.

Based on preparations of the sentence processing, the keyword extraction method of opinion sentences is proposed. The method input is n sentences provided by experts, and they are denoted as $S = \{s_1, s_2, \dots, s_n\}$, $i = 1, 2, \dots, n$. The output is the triples of keywords $EK_i = \{f_i, z_i, p_i\}$, $i = 1, 2, \dots, n$. The concrete procedure is as follows:

- (1) The i -th sentence s_i is segmented into k_i words with FMM.
- (2) The k_i words are tagged the part of speech with the Viterbi algorithm based on "PFR 'The People's Daily' Tagged Corpus".
- (3) The k_i words tagged with the part of speech are matched in the stop word list, and the useless words are filtered. Then the number of words is k'_i .
- (4) Based on "Chinese thesaurus (extended)" of HIT, words about alternative approaches form the set F , words about decision criterions form the set Z , and words about assessments form the set P . The semantic similarity between k'_i words and all the words in sets F, Z, P is denoted as $Sim_{j,t(r)}$, $j = 1, 2, \dots, k'_i$, $t(r)$ is the number of words in sets F ($r = 1$), Z ($r = 2$), P ($r = 3$).
- (5) The threshold of similarity is set as Th . The j -th word is placed in the r -th position of the triples EK_i if $Sim_{j,t(r)} > Th$.
- (6) All the sentences are processed with the above steps, and all the triples of keywords are obtained.

2.3.2. Decision Matrix Formation

The decision matrix is the basic of rational decision-making methods. It gathers the decision elements into a mathematical expression, including alternative approaches, decision criterions and assessment values. In the expert decision support mechanism proposed in this paper, the matrix components come from the processing results of experts' opinions. Assume that A is the alternative approach set and $A = \{a_1, a_2, \dots, a_m\}$, $i = 1, 2, \dots, m$. C is the decision criterion set and $C = \{c_1, c_2, \dots, c_n\}$, $j = 1, 2, \dots, n$. E is the expert set and $E = \{e_1, e_2, \dots, e_p\}$, $k = 1, 2, \dots, p$. Instead of one expert, several experts are chosen to increase the diversity of decision opinions. r_{ij}^k is the satisfaction assessment of the i -th alternative approach to the j -th decision criterion given by the k -th expert. The form of decision matrix is as follows:

$$R^k = \begin{pmatrix} r_{11}^k & \dots & r_{n1}^k \\ \vdots & \ddots & \vdots \\ r_{m1}^k & \dots & r_{mn}^k \end{pmatrix} \tag{6}$$

In decision matrix R^k , the components are from the triples of keywords EK in Section 2.3.1. Concretely, the alternative approaches in A consist of the elements f and the decision criterions in C consist of the elements z in triples EK . A tetrad of expert opinions is formed by adding the expert information into the triples of keywords. The tetrad is denoted as $r^k = \{exp^k, f_t^k, z_t^k, p_t^k\}$, where exp^k is the serial number of experts, f_t^k, z_t^k and p_t^k correspond to the alternative approach, decision criterion and assessment in the t -th sentence of the k -th expert, respectively. Based on the tetrads, a matrix-filling algorithm is proposed to connect the tetrads with the decision matrix. The concrete flow of the algorithm is shown in Figure 3.

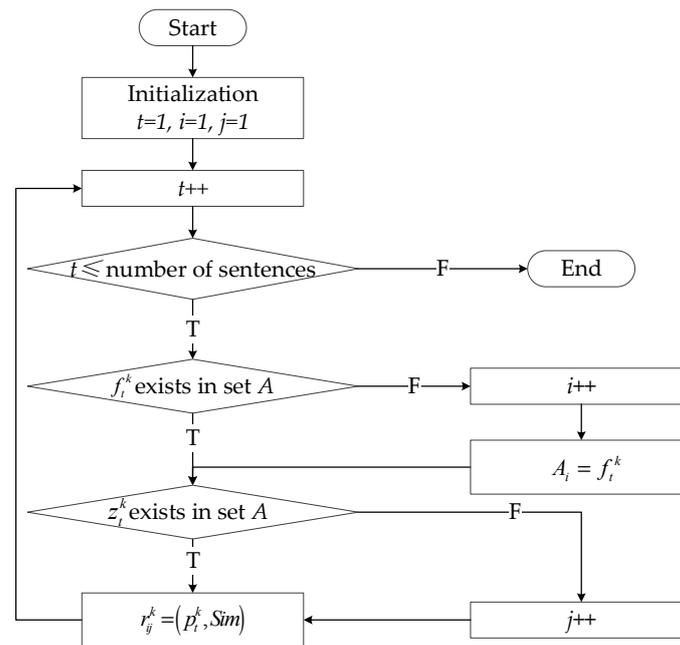


Figure 3. The flow of the matrix-filling algorithm which can determine the components of the matrix from the tetrads.

Each tetrad of an expert can form a decision matrix R^k . In the step of initialization, the alternative approach $a_i = f_i^k$, the decision criterion $c_j = z_j^k$, and $r_{ij}^k = (p_i^k, Sim)$, where *Sim* means the similarity of assessment value and fixed levels.

2.3.3. Multi-criteria Group Decision-Making

In the decision matrix, each element is expressed with a linguistic variable (p_i^k) and a similarity value (*Sim*), of which the form is similar to the two-tuple linguistic variable [23,24]. Therefore, the multi-criteria group decision-making method is proposed based on the two-tuple linguistic variable.

In a two-tuple linguistic (s, a) , s is a linguistic variable which means assessment degree, and a is a real number which means the membership of evaluated object to assessment degree. The element in decision matrix $r_{ij}^k = (p_i^k, Sim)$ is denoted as $r_{ij}^k = (s_{ij}^k, a_{ij}^k)$ to be accordant with the two-tuple linguistic. The weights of decision criteria are set as $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$.

The two-tuple weighted harmonic averaging operator (TWHA) is introduced to aggregate the elements in rows of decision matrix. TWHA is defined as

$$(\tilde{s}^t, \tilde{a}^t) = TWHA_{\omega}((s_1, a_1), (s_2, a_2), \dots, (s_n, a_n)) = \Delta \left(1 / \sum_{j=1}^n \frac{\omega_j}{\Delta^{-1}(s_j, a_j)} \right) \quad (7)$$

where $\{(s_1, a_1), (s_2, a_2), \dots, (s_n, a_n)\}$ is a group of two-tuple linguistic variables, and its weight vector is $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$, $\omega_j \in [0, 1]$ and $\sum_{j=1}^n \omega_j = 1$.

For matrix R^k , elements of the i -th row are aggregated with TWHA operator, and the integrated two-tuple of alternative a_i from expert e_k is obtained:

$$z_i^k = (s_i^k, a_i^k) = TWHA_{\omega}((s_{i1}^k, a_{i1}^k), (s_{i2}^k, a_{i2}^k), \dots, (s_{im}^k, a_{im}^k)) = \Delta \left(1 / \sum_{j=1}^m \frac{\omega_j}{\Delta^{-1}(s_{ij}^k, a_{ij}^k)} \right) \quad (8)$$

The two-tuple combined weighted harmonic averaging operator (TCWHA) is introduced to aggregate the assessments from different experts. TCWHA is defined as

$$(\hat{s}, \hat{a}) = TCWHA_{\omega,w}((s_1, a_1), (s_2, a_2), \dots, (s_n, a_n)) = \Delta \left(1 / \sum_{j=1}^n \frac{w_j}{\Delta^{-1}(\hat{s}_j, \hat{a}_j)} \right) \tag{9}$$

where (\hat{s}_j, \hat{a}_j) is the j -th largest element of the descending ranking result of the data group $\{\Delta(\Delta^{-1}(s_1, a_1)/n\omega_1), \Delta(\Delta^{-1}(s_2, a_2)/n\omega_2), \dots, \Delta(\Delta^{-1}(s_n, a_n)/n\omega_n)\}$, in which w_j is the reciprocal of ω_j and n is the balance factor.

For the alternative approach a_i , the assessments from different experts are aggregated with TCWHA operator, and the final assessment of a_i is obtained:

$$z_i = (s_i, a_i) = TCWHA_{\omega,w}((s_i^1, a_i^1), (s_i^2, a_i^2), \dots, (s_i^p, a_i^p)) = \Delta \left(1 / \sum_{j=1}^m \frac{w_j}{\Delta^{-1}(\hat{s}_j^k, \hat{a}_j^k)} \right) \tag{10}$$

The final result z_i indicates the assessment of alternative a_i from different decision criterions by several experts. Alternative approaches can be ranked by comparing the result z_i .

3. Experiment and Result

Algal bloom has become the typical pollution in urban lakes, which destroys the landscape and threatens drinking water safety. We choose Lake Yuyuantan in Beijing as the experiment object. The situation is set that algal bloom is to break out in the prediction based on water quality monitoring data. The expert decision support technique is conducted to select the suitable governance approaches to algal bloom in advance.

The experiment procedures include the choice of experts based on the classification, the processing of experts' opinions and the decision-making calculation.

3.1. Expert Classification and Selection

The experts' information is collected from China National Knowledge Internet (CNKI) and their affiliation's official websites. In the experiment, 22 experts are selected to form the basic expert database, and the expert list is shown in Table A1 of Appendix A. The experts' information includes their papers published and brief introductions about the professional title, academic degree and research field.

Following the classification method in Section 2.2, the to-be-classified object set is $V = \{v_1, v_2, \dots, v_{22}\}$, and each expert corresponds to a sentence set Y_i . Each text Y_i is segmented with the FMM algorithm. All the elements in the word segmentation set DW_i are scanned and matched with keywords in KW . The default keyword set $KW = \{kw_1, kw_2, \dots, kw_9\} = \{\text{professor, associate professor, doctor, master, reservoir and lake, Lake Taihu, algal bloom, sewage, urban}\}$. Then the attribute matrix of the to-be-classified objects M_{att} is formed, as shown in Table 1, in which values of matrix elements mean the repetition times of keywords.

Table 1. Attribute matrix of the to-be-classified objects M_{att} .

Expert	kw_1	kw_2	kw_3	kw_4	kw_5	kw_6	kw_7	kw_8	kw_9
Exp1	1	0	1	0	0	20	3	0	0
Exp2	1	0	1	0	6	0	1	0	1
Exp3	1	0	1	0	0	0	18	1	3
Exp4	1	0	1	0	0	0	12	0	1
.....
Exp11	0	1	1	0	0	0	3	0	0

In the applying of SOM to classify experts, the experts in M_{att} are divided into two groups: the first 11 experts are set as training samples and the others are testing samples. They are input to SOM and the matching degree S is calculated with the assumption that the anticipated classification is $\{Exp3, Exp4; Exp1, Exp2; Exp5; Exp6, Exp7, Exp8, Exp9, Exp10, Exp11\}$. The training result is shown in Table 2.

Table 2. Expert classification of 11 training samples.

Training Times	Class 1	Class 2	Class 3	Class 4	Matching Degree S
10	<i>Exp3, Exp4</i>		<i>Exp6, Exp5, Exp7, Exp8, Exp11</i>	<i>Exp2, Exp1, Exp9, Exp10</i>	7/3
30	<i>Exp3, Exp4</i>	<i>Exp1</i>	<i>Exp5</i>	<i>Exp2, Exp6, Exp7, Exp8, Exp9, Exp10, Exp11</i>	7/2
50	<i>Exp3, Exp4</i>	<i>Exp1, Exp2</i>	<i>Exp5</i>	<i>Exp10, Exp6, Exp7, Exp8, Exp9, Exp11</i>	MAX
100	<i>Exp3, Exp4</i>	<i>Exp1</i>		<i>Exp2, Exp6, Exp5, Exp7, Exp8, Exp9, Exp10, Exp11</i>	5/2

The best network (50 training times in Table 2) is selected by judging the matching degree S . Then the rest of the testing samples are input to the selected network, and the classification result is shown in Table 3.

Table 3. Expert classification of 11 testing samples.

Class 1	Class 2	Class 3	Class 4
<i>Exp12</i>	<i>Exp16, Exp19</i>	<i>Exp18, Exp20</i>	<i>Exp17, Exp13, Exp22, Exp14, Exp21, Exp15</i>

The classification results show that the experts in Class 1 are all professors and doctors who specialize in the urban lakes, while natural large-scale lakes are research objects for experts in Class 2. Experts in Class 3 focus on the algae in a microcosmic view with a weak relation with actual lakes. Experts in Class 4 are almost masters whose papers are fewer than those in other classes. The classification results reflect experts' educational background, major field and research level. Suitable experts can be chosen from different classes with specific demands.

Considering the experiment background and expert classification results, three experts who specialize in urban algal bloom are chosen to provide decision support opinions. The experts are *Exp3*, *Exp4* and *Exp12* in Class 1.

3.2. Decision Opinion Processing

The three experts are invited to provide decision opinions based on water quality monitoring data in real time. For the convenient calculation in computers, the guidance expression form of opinions is given that each opinion should be a sentence which may include alternative approach, decision criterion and assessment level.

The keyword extraction of original opinions is conducted following the methods in Section 2.3.1, including the word segmentation, part-of-speech tagging, word filtering and semantic similarity computation. The triples of keywords are shown in Table 4.

Table 4. Triples of keywords extracted from expert opinions.

Expert	Triples of Keywords from Opinions
e_1	{mechanical removal, investment scale, lower} {mechanical removal, speed, fast} {microorganism, speed, slow}
e_2	{aeration, investment scale, very high} {microorganism, environment influence, lower}
e_3	{mechanical removal, environment influence, low} {electrochemistry, investment scale, very much}

Based on the keyword extraction, decision matrix can be formed following the method in Section 2.3.2. According to the keywords, alternative approaches and decision criteria are confirmed. $A = \{\text{mechanical removal, aeration, flocculate precipitation, electrochemistry, algaecide, microorganism}\}$, $C = \{\text{investment scale, removal degree, influence on environment, efficiency, sustainable development}\}$. The assessment grades are set as five levels, including very high (VH), high (H), medium (M), low (L) and very low (VL).

The triples of keywords are transformed to the tetrads based on which decision matrix is filled, as shown in Table 5.

Table 5. Decision matrix from three experts’ opinions.

		c_1	c_2	c_3	c_4	c_5
e_1	a_1	(L, 0.55)	(VH, 0.11)	(L, 0.36)	(VH, 0.83)	(L, 0.09)
	a_2	(H, 0.10)	(L, 0.81)	(L, 0.13)	(H, 0.75)	(L, 0.47)
	a_3	(M, 0.68)	(L, 0.34)	(L, 0.91)	(L, 0.19)	(L, 0.80)
	a_4	(VH, 0.93)	(L, 0.25)	(H, 0.18)	(L, 0.01)	(VL, 0.12)
	a_5	(H, 0.60)	(H, 0.63)	(VH, 0.67)	(M, 0.87)	(VL, 0.32)
	a_6	(H, 0.71)	(M, 0.62)	(VL, 0.00)	(VL, 0.26)	(H, 0.20)
e_2	a_1	(M, 0.32)	(H, 0.67)	(VL, 0.06)	(H, 0.93)	(VL, 0.75)
	a_2	(VH, 0.22)	(M, 0.11)	(L, 0.64)	(VH, 0.13)	(VL, 0.82)
	a_3	(M, 0.95)	(L, 0.57)	(L, 0.62)	(L, 0.70)	(VL, 0.32)
	a_4	(H, 0.49)	(VL, 0.95)	(VH, 0.75)	(L, 0.60)	(L, 0.21)
	a_5	(M, 0.13)	(M, 0.22)	(VH, 0.04)	(H, 0.82)	(L, 0.21)
	a_6	(H, 0.01)	(VL, 0.96)	(L, 0.13)	(VL, 0.83)	(H, 0.16)
e_3	a_1	(VL, 0.79)	(VH, 0.02)	(L, 0.26)	(VH, 0.31)	(L, 0.58)
	a_2	(M, 0.95)	(VL, 0.27)	(M, 0.71)	(H, 0.15)	(M, 0.53)
	a_3	(M, 0.54)	(L, 0.65)	(L, 0.70)	(M, 0.26)	(M, 0.34)
	a_4	(VH, 0.24)	(L, 0.64)	(M, 0.38)	(L, 0.20)	(L, 0.15)
	a_5	(H, 0.69)	(M, 0.98)	(VH, 0.29)	(H, 0.88)	(L, 0.86)
	a_6	(H, 0.14)	(M, 0.09)	(VL, 0.75)	(L, 0.20)	(VH, 0.46)

3.3. Decision-Making Calculation

For the decision matrix in Table 5, the multi-criteria group decision-making method is conducted. The i -th row elements of the matrix are aggregated with TWHA operator (following Formulas (7) and (8)). The evaluation of alternative approaches is obtained from five decision criteria. Then different experts’ assessments are aggregated with TCWHA operator (following Formulas (9) and (10)). Table 6 shows the aggregated evaluation of different experts and the final evaluation for alternative approaches. The comparison of alternative approaches is shown in Figure 4.

The sorting of alternative approaches is $a_1 \succ a_3 \succ a_6 \succ a_2 \succ a_5 \succ a_4$, and it accords with the limitation of decision criteria and actual comprehension. For example, the best approach a_1 (mechanical removal) is strongly high in the governance efficiency and removal degree, and it is low in the environmental influence and investment scale; while a_5 (algaecide) and a_4 (electrochemistry) have negative influence on the environment and they display badly in the removal efficiency and investment scale.

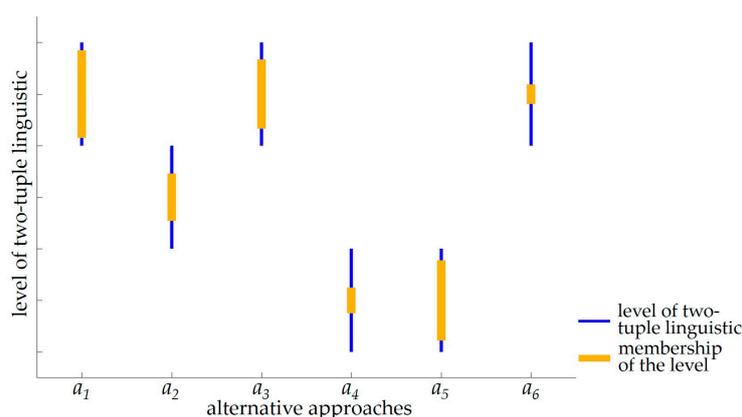


Figure 4. Comparison of alternative approaches. For different levels of two-tuple linguistic, the higher the level, the better the alternative approach. For the same level, the larger the membership, the better the alternative approach.

Table 6. Evaluation for alternative approaches in aggregation for each expert and final result.

	a_1	a_2	a_3	a_4	a_5	a_6
e_1	(M, 0.95)	(VL, 0.27)	(M, 0.71)	(VH, 0.75)	(L, 0.60)	(L, 0.21)
e_2	(M, 0.54)	(L, 0.65)	(L, 0.70)	(VH, 0.04)	(H, 0.82)	(L, 0.21)
e_3	(VH, 0.24)	(L, 0.64)	(M, 0.38)	(L, 0.81)	(L, 0.13)	(H, 0.75)
final result	(H, 0.85)	(M, 0.46)	(H, 0.67)	(L, 0.25)	(L, 0.78)	(H, 0.19)

4. Discussion

An expert decision support technique is proposed to explore the solution to algal bloom pollution in urban lakes. The decision support mechanism is set as the general framework to conduct the decision-making based on text analysis. The text processing and decision-making calculation technique are the core methodology in this paper.

In the traditional management of algal bloom pollution, governance approaches are selected relying on administrators’ personal experience and knowledge. The selection lacks a clearly specified procedure which can be supplemented by the decision support mechanism. In the mechanism, the expert database is formed to gather expert resources, and the processing technique is proposed to transform and calculate professional decision opinions. The decision support mechanism focuses on two aspects: experts’ support and standard procedure. In terms of experts’ support, experts evaluate governance approaches with professional opinions, which supplement the specialized knowledge shortage of administrators. Moreover, in the decision procedure, experts can receive quantitative information to analyze the issue with real-time monitoring data, predictive trend information and climate data. Both the professional cultivation and real-time comprehensive analysis can optimize the decision procedure. In terms of standard procedure, decision support mechanism consists of definite steps which are all presupposed and executable. The standard and computable procedure makes it possible to realize the information processing and decision-making calculation automatically and rapidly within an information system.

Water pollution management faces the challenge and opportunity of big data and the information era. Techniques in interdisciplinary fields have been applied in water environment research [17,25]. The decision-making technique based on text analysis provides a new approach to algal bloom governance with methods in management science and computer science. It performs further on the decision-making than the previous work. In work [13–15], the decision-making is based on evaluated scores of alternatives from different aspects. Alternatives are assessed by synthesizing scores with methods such as Bayes and fuzzy multi-attribute. In work [16,17], the evaluation expression is improved with Vague set, whose original form is verbal rating scale. The simple form in previous work

may distort or drop the natural meaning of original decision opinions. We choose natural language to enhance the objectivity and integrity of opinions. The processing and calculation of text information make an effort to retain and reflect the natural meaning. Besides the opinion expression form, we try to dynamically adjust the decision matrix. This paper confirms the components of decision matrix under the analysis results of experts' decision opinions, while the previous work determines decision criteria and alternatives in the precondition. The dynamic formation of decision matrix makes decision-making adaptive to concrete situations.

This paper broadens the solution to the decision-making of algal bloom governance. While it explores the solution, more work needs to be done to perfect the expert decision support technique. For text processing, novel techniques in the artificial intelligence can be applied to mine texts more deeply and rapidly. Instead of using monitoring data indirectly, experts' opinions should be fused with monitoring data to support the decision-making procedure.

5. Conclusions

For algal bloom governance in urban lakes, an expert decision support technique is proposed in this paper, including the decision mechanism from the management aspect and the processing calculation methods from the realization aspect.

In the decision mechanism, experts are brought in to enhance the professional level. Experts' specific research achievements can supply the knowledge shortage of algal bloom governance for public administrators. In the processing calculation, several methods are introduced to analyze experts and their decision opinions.

The expert decision support technique strengthens the utilization of Internet information and realizes the automatic data processing to a certain extent. It is an exploration to text analysis in the decision-making for algal bloom governance in urban lakes and it can be expanded to the similar water pollution management in the big data era. The expert decision support technique can be developed more widely and deeply to assist in the protection of urban water environment.

Acknowledgments: This work has been supported by the Postgraduate Training Joint-Build Project of Beijing Municipal Commission of Education (No. 20120639002), National Natural Science Foundation of China (No. 61573061), National Natural Science Foundation of China (No. 51179002) and National Natural Science Foundation of China (No. 61673002).

Author Contributions: Yu-Ting Bai conceived and proposed the decision support technique, conducted the experiments and wrote the paper; Bai-Hai Zhang and Xiao-Yi Wang modified the decision-making mechanism and wrote the related sections; Xue-Bo Jin designed the experiments; Ji-Ping Xu and Zhao-Yang Wang conducted the experiment of text processing.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The decision technique in this paper is based on the text processing. Decision texts come from the experts who are selected from the expert database. The member list of basic expert database in the Section *Experiment and Result* is shown in Table A1, in which the experts are numbered from *Exp1* to *Exp22*.

Table A1. Member list of basic expert database.

No.	Expert Name	No.	Expert Name	No.	Expert Name
<i>Exp1</i>	Kong Fanxiang	<i>Exp9</i>	Yang Bin	<i>Exp17</i>	Dong Zengchuan
<i>Exp2</i>	Du Guisen	<i>Exp10</i>	Ma Xinyu	<i>Exp18</i>	Pei Hongping
<i>Exp3</i>	Liu Zaiwen	<i>Exp11</i>	Wang Li	<i>Exp19</i>	Yang Liuyan
<i>Exp4</i>	Wang Xiaoyi	<i>Exp12</i>	Cui Lifeng	<i>Exp20</i>	Shi Yan
<i>Exp5</i>	Gao Yurong	<i>Exp13</i>	Yang Zhifeng	<i>Exp21</i>	Lv Siying
<i>Exp6</i>	Zhou Yunlong	<i>Exp14</i>	Dong Shuoqi	<i>Exp22</i>	You Liang
<i>Exp7</i>	Tang Lina	<i>Exp15</i>	Li Mengxun		
<i>Exp8</i>	Zhu Shiping	<i>Exp16</i>	Wu Tingfeng		

References

- Honti, M. Controlling river eutrophication under conflicts of interests—A GIS modeling approach. *Water* **2015**, *7*, 5078–5090. [CrossRef]
- Wang, X.Y.; Yao, J.Y.; Shi, Y.; Su, T.L.; Wang, L.; Xu, J.P. Research on hybrid mechanism modeling of algal bloom formation in urban lakes and reservoirs. *Ecol. Model.* **2016**, *332*, 67–73.
- China Environmental State Bulletin of 2015. Available online: http://www.zhb.gov.cn/gkml/hbb/qt/201606/t20160602_353138.htm (accessed on 2 June 2016).
- Beijing Environmental State Bulletin of 2015. Available online: <http://www.bjepb.gov.cn/bjepb/413526/413663/413763/413975/4387663/index.html> (accessed on 13 April 2016).
- Wang, X.Y.; Tang, L.N.; Liu, Z.W.; Xu, J.P. Research on the fuzzy petri net optimization modeling of water bloom formation process. *Acta Electron. Sin.* **2013**, *41*, 68–71.
- Zhang, Z.; Peng, G.Q.; Guo, F.; Hu, D.; Huang, C.C.; Song, Z.Y. The key technologies for eutrophication simulation and algal bloom prediction in lake Taihu, China. *Environ. Earth Sci.* **2016**, *75*, 1295. [CrossRef]
- Algal Bloom Broke Out Firstly in Chuhe River in 5 Years. Available online: <http://news.163.com/16/0809/16/BU1QNTJ900014AEE.html> (accessed on 9 August 2016).
- Liu, Z.W.; Wang, X.Y.; Cui, L.F. *Monitoring and Evaluation of Water Environment, Intelligent Prediction Methods and Emergency Governance Decision System on Algal Bloom*; Chemical Industry Press: Beijing, China, 2013; pp. 3–32.
- Li, Q.; Hu, W.; Zhai, S. Integrative indicator for assessing the alert levels of algal bloom in lakes: Lake Taihu as a case study. *Environ. Manag.* **2016**, *57*, 237–250. [CrossRef] [PubMed]
- Wang, S.B.; Xu, Z.R.; Zhang, J. A review of technologies for prevention and control of cyanobacteria blooms in large-scale eutrophicated lakes and reservoirs. *Water Resour. Prot.* **2016**, *32*, 88–99.
- Jia, Y.H.; Yang, Z.; Su, W.; Dan, J.; Kong, F.X. Controlling of cyanobacteria bloom during bottleneck stages of algal cycling in shallow Lake Taihu (China). *J. Freshw. Ecol.* **2014**, *29*, 129–140. [CrossRef]
- Waajen, G.W.A.M.; Bruggen, N.C.B.; Pires, L.M.D.; Lengkeek, W.; Lurling, M. Biomanipulation with quagga mussels (*Dreissena rostriformis bugensis*) to control harmful algal blooms in eutrophic urban ponds. *Ecol. Eng.* **2016**, *90*, 141–150. [CrossRef]
- Li, D.G. Bloom Research on Algal Bloom Forecast and Emergency Decision-Making Intelligence Method. Master's Thesis, Beijing Technology and Business University, Beijing, China, June 2011.
- Wang, X.Y.; Chen, C.; Liu, Z.W.; Xu, J.P. Research on the emergency control decision on water bloom in lake and reservoir based on fuzzy Bayes under comprehensive restrictions. *Intell. Syst. Design Eng. Appl.* **2012**, 894–898. [CrossRef]
- Liu, Z.W.; Li, L.; Wang, X.Y. Researches of water bloom emergency management decision making method and system based on fuzzy multiple attribute decision making. *Int. J. Comput. Sci. Issues* **2012**, *9*, 48–53.
- Bai, Y.T.; Wang, X.Y.; Wang, L.; Xu, J.P.; Yu, J.B.; Shi, Y. The research of decision-making method for multi-objective in water bloom emergency governance based on vague set theory. *J. Comput. Inf. Syst.* **2014**, *10*, 2099–2106.
- Bai, Y.T.; Zhang, B.H.; Wang, X.Y.; Jin, X.B.; Xu, J.P.; Su, T.L.; Wang, Z.Y. A novel group decision-making method based on sensor data and fuzzy information. *Sensors* **2016**, *16*, 1799. [CrossRef] [PubMed]
- Hu, X.H. Application of maximum matching method in Chinese segmentation technology. *J. Anshan Norm. Univ.* **2008**, *10*, 42–45.
- Xu, Y.F.; Li, J.; Wei, Y.T. Clustering analysis on airport weather based on SOM network. *Math. Pract. Theory* **2016**, *46*, 210–217.
- The People's Daily Tagged Corpus. Available online: http://www.icl.pku.edu.cn/icl_res (accessed on 26 December 2016).
- Zhang, W. Research and Realization of Part-of-Speech Tagging in Chinese. Master's Thesis, Nanjing Normal University, Nanjing, China, May 2007.
- Harbin Institute of Technology Chinese Thesaurus (Extended). Available online: http://ir.hit.edu.cn/demo/ftp/Sharing_Plan.htm (accessed on 26 December 2016).
- Herrera, F.; Martinez, L. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Trans. Fuzzy Syst.* **2000**, *8*, 746–752.

24. Beg, I.; Rashid, T. An intuitionistic 2-tuple linguistic information model and aggregation operators. *Int. J. Intell. Syst.* **2016**, *31*, 569–592. [[CrossRef](#)]
25. Luo, P.P.; He, B.; Chaffe, P.L.B.; Nover, D.; Takara, K.; Rozainy, M.R. Statistical analysis and estimation of annual suspended sediment of major rivers in Japan. *Environ. Sci. Process. Impacts* **2013**, *15*, 1052–1061. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).