

Expert Clustering of Hydraulic-Geotechnical Control Parameters of Rubble Mound Breakwaters Using Rough Sets Theory

GH. A. MONTAZER*

Assist. Prof. of Eng. Sch.

REZA SABZEVARI**

MSc. Student of Mechatronics Eng.

H. GH. POUR KHATIR

MSc. Student of Eng. Sch.

F. NASIRI SALEH

Ph.D. Student of Eng. Sch.

*School of Engineering, Tarbiat Modarres University

Address: P. O. Box 14115-179, Tehran, Iran; Tel: (021) 88011001- 3990

**School of Engineering, Department of Mechatronics Engineering

Islamic Azad University of Qazvin, Qazvin, Iran

Member of Young Researchers' Club (YRC)

Abstract: - The problem of how various parameters of sea and soil affect the hydraulic and geotechnical aspects of breakwaters and the way these parameters must be controlled stably despite of undesired environmental factors is a challenging task in Hydraulics Engineering. The most popular solution for this problem is based on forming an information table whose entities have been extracted from empirical experiments and then try to find the relations between the affecting parameters and affected elements. In this paper, a new expert method based on the notions of Rough Sets Theory (RST) has been developed to solve this problem. Also, this development method has been applied to the problem of clustering the hydraulic-geotechnical control parameters of a real rubble mound breakwater and then the results and advantages of this approach have been explained and discussed.

Key-Words: - Intelligent approach, Hydraulic design, Clustering, Breakwater, Rough Sets Theory

1 Introduction

Design of breakwaters is one of the most important problems in coastal engineering. Rubble mound breakwater, one of the types of breakwaters, has been used widely in Iran because of its simplicities in design and implementation [1]. For designing the rubble mound breakwater, the breakwater's cross-section should be calculated and also the dimension of Armour units should be determined. To design breakwater's cross-section, experimental methods are used and thereby parameters like thickness of primary (Armour), secondary Armour and filter layers, width and level of breakwater's crown, level of core's crown, and slope and toe of breakwater are determined. Finally, for designing the dimension of Armour units, applying experimental methods and formulas the weight of required Armour is determined [3].

After hydraulic design of breakwaters, their geotechnical stability is controlled using numerical or

analytical methods. The most popular solution for this problem is based on forming an information table whose entities have been extracted from empirical experiments and then try to find the relations between the affecting parameters and affected elements, which have been named as Hydro-Informatics. Many approaches such as Dimensional Analysis [4] and Artificial Neural Networks [8] have been applied to design and control the structure of breakwaters, but all of them have to use data from an information table, which is often vague and probably uncertain due to applied experimental methods and measurements. In this paper, a new expert method based on the notions of Rough Sets Theory (RST) has been developed to solve this problem. This approach is a powerful tool for data analysis, which has attracted attentions of many researchers all over the world [9, 5, 13 and 6]. In this article, the Rough Sets Theory has been applied to the problem of clustering hydraulic-geotechnical control parameters of a real rubble mound breakwater whose database has affecting elements such as [1]:

Wave significant height (H_s), Wave period (T_s), Density of Armour (ρ_s), Density of sea water (ρ_w), Breakwater permeability (p), Number of incidence waves (N), Breakwater slope ($\cot(\alpha)$), Damage level (S), Depth of water (d_p), Minimum depth of water (d_{min}), Breakwater friction angle (ϕ_{BRE}), Bed friction angle (ϕ_b), Density of bed materials (ρ_{sb}), and two affected elements: Nominal diameter of Armour layer (D_{n50}) and Stability factor (SF).

This database has more than 1400 rows whose values have been determined by empirical experiments or software simulations [1]. The proposing RST approach does an expert analysis on the entities of this database, extracts the hidden relations between any two-table entities, and then defines the functionality of the affecting and affected control parameters in hydraulic-geotechnical design of breakwaters.

2 Basic Concepts of the Rough Sets Theory

Rough Sets Theory (RST) was introduced by Pawlak in 1982 [10]. It has attracted the attention of many researcher and practitioners all over the world, who contributed to its development and application during the last decade [11, 9, 12 and 7]. This theory describes dependencies between attributes, to evaluate significance of attributes, and to deal with inconsistent data.

The Rough Sets philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (i.e. data knowledge). Objects characterized by the same information are indiscernible in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis for the Rough Sets Theory [9].

By an information system S , we mean $S = \{U, Q, V, f\}$, where $U = \{x_1, x_2, \dots, x_n\}$ is a finite set of object, $Q = \{q_1, q_2, \dots, q_m\}$ is a finite set of attributes which this set is further classified into two disjoint subsets, condition attribute C , and decision attribute D , $Q = C \cup D$ and V_p is the domain of values for attribute p such as:

$$V = \cup \{V_p \mid p \in Q\} \tag{1}$$

and $f: U \times Q \rightarrow V$ is a total function such that $f(x_i, q) \in V_q$ for every $q \in Q, x_i \in U$.

Let $E \subseteq Q$ and $x, y \in U$. We say that x, y are indiscernible by the set of attribute E in S , if $f(x, q) = f(y, q)$ for every $q \in E$. Thus every $E \subseteq Q$ generates a

binary relation denoted by I_E . Obviously, I_E is an equivalence relation for any E . Equivalence classes of the relation I_E are called E -elementary sets in S and $I_E(x)$ denotes the E -elementary set containing the object $x \in U$, hence [7]:

$$I_E(x) = \{y \in U \mid \forall q \in E, f(x, q) = f(y, q)\} \tag{2}$$

Now let $E \subseteq Q$ and $Y \subseteq U$. The E -Lower approximation of Y , denoted by $\underline{E}Y$, and the E -Upper approximation of Y , denoted by $\bar{E}Y$ are defined respectively as:

$$\bullet \underline{E}Y = \cup \{x \in U \mid I_E(x) \subseteq Y\} \tag{3}$$

$$\bullet \bar{E}Y = \cup \{x \in U \mid I_E(x) \cap Y \neq \Phi\} \tag{4}$$

$$\bullet \text{The } E\text{-boundary of set } Y \text{ is defined as } \text{BND}(Y) = \bar{E}Y - \underline{E}Y. \tag{5}$$

Set $\underline{E}Y$ is the set of all elements of U , which can be certainly classified as elementary of Y , employing the set of attribute E . Set $\bar{E}Y$ is the set of elements of U , which can be possibly classified as elements of Y using the set of attribute E . The set $\text{BND}(Y)$ is the set of elements, which cannot be certainly classified as elements of Y , using the set of attribute E .

3 The New Methodology for Rough Sets Based Solving Approach

Consider the example of an information table which has been shown in table 1. In this knowledge system, we have:

- The set of objects $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$,
- The set of condition attributes $C = \{c_1, c_2, c_3, c_4\}$,
- The set of decision attributes $D = \{d\}$,
- Particular sets of attribute values given as: $V_{c_1} = \{A, B\}$, $V_{c_2} = \{\text{Yes}, \text{No}\}$, $V_{c_3} = \{10, 50\}$, $V_{c_4} = \{-5, 7\}$, $V_d = \{\text{Low}, \text{High}\}$.
- The set of attribute values is given as: $V = \{V_{c_1}, V_{c_2}, V_{c_3}, V_{c_4}, V_d\}$.

Table 1. An information system.

U	Q				
	C				D
	c ₁	c ₂	c ₃	c ₄	d
x ₁	A	Yes	10	-5	Low
x ₂	B	Yes	10	-5	High
x ₃	A	Yes	10	7	High
x ₄	A	Yes	50	7	High
x ₅	B	No	10	-5	High
x ₆	A	No	10	-5	Low

Let $Y = \{x \in U \mid d(x) = \text{High}\} = \{x_2, x_3, x_4, x_5\}$ and $E = \{c_2, c_3\}$, Then:

- $I_E = \{\{x_1, x_2, x_3\}, \{x_4\}, \{x_5, x_6\}\}$.
- $I_E(x_1) = \{x_1, x_2, x_3\}$
- $I_E(x_5) = \{x_5, x_6\}$
- $\underline{E}Y = \{x_4\}$
- $\bar{E}Y = \{x_1, x_2, x_3, x_4, x_5, x_6\}$,
- $\text{BND}(Y) = \{x_1, x_2, x_3, x_5, x_6\}$.

Algorithms used for classification of a decision tables can be presented as algebraic developments or logical relations. Many algorithms have been developed to reduce the conditions and have been used in many problems [5 and 13]. In this paper we presented a modified procedure, collecting useful parts of previous approaches. The results of this procedure are return by a java program. This program uses an input file to analyze the data and then put the output data into another file. Basic steps in data analysis, which can be tackled employing this program, are as following:

Step 1: The first step of this algorithm is to eliminate unnecessary input variables from the table. This task can be accomplished eliminating each attribute and verifying if lower approximation of the resulting table is equal to lower approximation of the original one. In table 1 we have $\underline{C}U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$. If we eliminate c_1 then $\underline{P}_1 = \{c_2, c_3, c_4\}$ and $\underline{P}_1U = \{x_3, x_4\}$. Therefore because of $\underline{C}U \neq \underline{P}_1U$, we cannot eliminate c_1 . If we examine eliminating c_2 and c_3 , we can show that $\underline{P}_2 U = \underline{C}U$ and $\underline{P}_3 U = \underline{C}U$. Therefore, we can eliminate c_2 and c_3 . But if we examine eliminating c_4 , we have $\underline{P}_4 \neq \underline{C}$, so c_4 could not be eliminated. Hence, using this step the table 1 can be reduced to the table 2.

Table 2. Reduced information system by step 1.

U	Q		
	C		D
	c ₁	c ₄	d
x ₁	A	-5	Low
x ₂	B	-5	High
x ₃	A	7	High
x ₄	A	7	High
x ₅	B	-5	High
x ₆	A	-5	Low

Step 2: The second step is to eliminate objects that are repeated in table 2. Using this step the resulting table is shown in table 3.

Table 3. Reduced information system by step 2.

U	Q		
	C		D
	c ₁	c ₄	d
x ₁	A	-5	Low
x ₂	B	-5	High
x ₃	A	7	High

Step 3: The third step is to remove unnecessary value of attributes for each decision rule. This is known as finding the *core* values. This task can be accomplished eliminating each condition attribute value and verifying whether the table is consistent. A table is consistent if for every combination of condition attributes presented in the table 3, a unique value for the decision attribute could be achieved. In table 3, If we eliminate the value of $(x_1, c_1) = A$, then the table become inconsistent. Therefore, we cannot eliminate this value. But we can eliminate the values of $(x_3, c_1) = A$, $(x_2, c_2) = -5$. Applying this step the resulting table has been shown in table 4.

Table 4. Reduced information system by step 3.

U	Q		
	C		D
	c ₁	c ₄	d
x ₁	A	-5	Low
x ₂	B	-	High
x ₃	-	7	High

Step 4: The next step is to eliminate objects that are repeated in the resulting table which in this example, table 4, there is no object that has been repeated.

Step 5: The last step is to express the resulting table some *IF ... THEN ...* statements:

- *IF* (c_1 is A and c_2 is -5) *THEN* (d is Low)
- *IF* (c_1 is B) *THEN* (d is High)
- *IF* (c_2 is 7) *THEN* (d is High)

4 Applying the New Method in Solving Hydraulic Engineering Problem

We have a database which has the abstract of expertness of specialists and the experimental results about hydraulic-geotechnical aspects of breakwater.

This database has composed of thirteen measured items, including the parameters of Wave significant height (H_s), Wave period (T_s), Density of Armour (ρ_s), Density of sea water (ρ_w), Breakwater permeability (P), Number of incidence water (N), Breakwater slope ($\cot(\alpha)$), Damage level (S), Depth of water (d_p), Minimum depth of water (d_{\min}), Breakwater friction angle (Φ_{BRE}), Bed friction angle (Φ_b), and Density of bed material (ρ_{sb}) as condition attributes (C) and the parameters of nominal diameter of Armour layer (D_{n50}) and stability factor (SF) as decision attributes (D). Our sample database has 1440 objects which show the empirical relation among condition attributes and decision ones. These data have been extracted from experimental measurement using MSTAB and BREAKWATER software [1]. Some of these database entities have been shown in table 5. Basic problems blinking in this database are the vagueness of decision variables and also the uncertain relation between object-attribute values and their corresponding results in decision columns. It is obvious that the larger size of database, the more difficulties in decision processes. In this paper, we use the method described above to reduce the size of foregoing information system and classify the data entities to ease the decision-making process. As the first step of our algorithm we eliminated every condition attributes of the table 5 and knew that for $D = \{D_{n50}\}$, 7 attributes (ρ_w , P, N, d_p , d_{\min} , Φ_{BRE} , Φ_b) and for $D = \{SF\}$, 4 attributes (ρ_w , P, N, d_p) could be omitted. Then using the second step to eliminate repeated objects, we find that for $D = \{D_{n50}\}$, 237 objects and for $D = \{SF\}$, 634 objects could be eliminated. In third step, we eliminated condition attribute values without losing the consistency of our table and at last in step 4, the repeated objects are omitted. Following the steps described above, two tables are resulted with totally 15 condition attributes and 3654 values for these attributes. In comparison with the original information table which has totally 26 condition attributes, for two decision ones, and 38880 different values for condition attributes, it is obvious that the rate of reduced data is very high. Some results have been shown in tables 6 and 7.

5 Conclusion

The knowledge acquisition process is the most difficult task during the construction of knowledge-

based systems, such as expert system and case-based reasoning. Usually, the experts have some difficulties in presenting the knowledge and the way they solve a given problem. This fact may result in superfluous information and/or an incomplete set of information.

This paper has presented a systematic approach for compacting the hydraulic-geotechnical database. This approach uses Rough Sets Theory (RST) and the concepts of core and reducts of knowledge. This method has been based on the logical concepts and uses arithmetic tools to reduce the size of database and results in core of knowledge. Applying this method to a practical problem of clustering control parameters of an especial rubble mound breakwater shows a dramatic reduction of data entities (from 26×38880 to 15×3654) which has deep influences on the simplicity of the processes and save much more time for engineers and computers. Also, this research shows that the number of effective parameters on D_{n50} is 6 and for SF is 9 rather than 13. This method can be generalized to all of similar problems and the developed package has the capability to solve other various problems.

References:

- [1] Ataei, A.; *Design of Rubble Mound Breakwaters Using Artificial Neural Networks*, M. Sc. Thesis, Tarbiat Modarres University, Tehran, Iran, Winter 2002.
- [2] Allsop, N.W.H. and Wood, L.A.; *Hydro-Geotechnical Performance of Rubble Mound Breakwater: Literature*, Report SR 98, Walling ford, Hydraulics Research, March 1987.
- [3] CUR, Center for Civil Engineering Research and Codes, *Manual on the use of Rock in Hydraulic Engineering*; A.A.Balkema / Rotterdam /Brookfield, 2000.
- [4] Godsian, M., N. Amanian and S. A. Marashi Shooshtari; *Discharge Coefficient of Semicircular Laby-Rinth Weir*; Amirkabir Journal of Science & Technology; Vol. 13, No. 49, Winter 2002, pp. 76-83.
- [5] Kusiak, A.; *Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing*; IEEE Transactions on Electronics Packaging Manufacturing; Vol. 24, No. 1; January 2001.
- [6] Montazer, Gh. A.; *A New Approach for Knowledge-Based Systems Reduction Using Rough Sets Theory*; Int'l J. of Eng.; Spring 2003.
- [7] Montazer, Gh. A. and Tayefeh, M. R. ; *A New Approach for Statistical Data Analysis Using Rough*

Sets Techniques; Proc. of the 6th Int'l Statistics Conf.; Tehran, Iran, Sep. 2002, pp. 146-150.

[8] Nasiri Saleh, F. and Gh. A. Montazer; *Determination of Scour Hole Dimensions Down Streams of Free Over Fall Spillway Using Artificial Neural Network*; Proc. of 3rd Iranian Conf. on Hydraulic Eng.; Tehran, Iran; 2001, pp. 301-308.

[9] Nowicki, R., R. Slowinski and J. Stefanowski; *Evaluation of Vibroacoustic Diagnostic Symptoms by Means of the Rough Sets Theory*; Computers in Industry; Vol. 20, No. 2; Aug. 1992; pp.141-152.

[10] Pawlak, Z.; *Rough Sets*; International Journal of Computer and Information Science; Vol. 11; 1982; pp. 341-356.

[11] Pawlak, Z; *Rough Sets, Theoretical Aspects of Reasoning about Data*; Kluwer Academic Publishers; London, 1992.

[12] Szladow, A. and W. Ziarko; *Rough Sets: Working with Imperfect Data*; AI Expert; Vol. 8; 1993; pp. 36-41.

[13] Yin, X., Z. Zhou, N. Li and Sh. Chen; *An Approach for Data Filtering Based on Rough Set Theory*; Wang X S, Yu G, Lu H eds. Lecture Notes in Computer Science 2118, Berlin: Springer-Verlag, 2001, pp. 367-374.

Table 5. Some of hydraulic database information

No.	Condition Attributes													Decision Attributes	
	H _s	T _s	ρ _s	ρ _w	P	N	Cot(α)	S	d _p	d _{min}	Φ _{BRE}	Φ _b	ρ _{sb}	D _{n50}	SF
1	3	8	2500	1025	0.4	5000	2	2	8	5	34	23	18	1.3744	1.234
2	3	8	2500	1025	0.4	5000	2	2	8	5	34	23	20	1.3744	1.291
3	3	8	2500	1025	0.4	5000	2	2	8	5	34	26	18	1.3744	1.331
4	3	8	2500	1025	0.4	5000	2	2	8	5	34	26	20	1.3744	1.378
5	3	8	2500	1025	0.4	5000	2	2	8	5	40	23	18	1.3744	1.331
6	3	8	2500	1025	0.4	5000	3	2	8	5	34	23	18	1.1222	1.444
7	3	8	2500	1025	0.4	5000	3	2	8	5	34	23	20	1.1222	1.525
8	3	8	2500	1025	0.4	5000	3	2	8	5	34	26	18	1.1222	1.592
9	3	8	2500	1025	0.4	5000	3	2	8	5	34	26	20	1.1222	1.669
10	3	8	2500	1025	0.4	5000	3	2	8	5	40	23	18	1.1222	1.54
11	3	11	2500	1025	0.4	5000	2	2	8	5	34	23	18	1.5384	1.241
12	3	11	2500	1025	0.4	5000	2	2	8	5	34	23	20	1.5384	1.301
13	3	11	2500	1025	0.4	5000	2	2	8	5	34	26	18	1.5384	1.47
14	3	11	2500	1025	0.4	5000	2	2	8	5	34	26	20	1.5384	1.394
15	3	11	2500	1025	0.4	5000	2	2	8	5	40	23	18	1.5384	1.346

Table 6. Some of reduced hydraulic database for D = {D_{n50}}

No.	H _s	T _s	ρ _s	Cot(α)	S	ρ _{sb}	D _{n50}
1	3	8	-	2	2	20	1.374
2	3	8	-	2	5	18	1.144
3	3	11	-	2	2	18	1.538
4	3	11	-	2	2	20	1.538
5	3	11	-	2	5	18	1.28
6	5	8	-	2	2	18	2.016
7	5	8	-	2	2	20	2.016
8	5	11	-	3	5	20	1.607
9	7	8	-	2	2	-	2.594
10	7	8	-	2	5	-	2.16

Table 7. Some of reduced hydraulic database for D = {SF}

No.	H _s	T _s	ρ _s	Cot(α)	S	d _{min}	Φ _{BRE}	Φ _b	ρ _{sb}	SF
1	3	8	-	2	2	-	34	26	18	1.331
2	3	8	-	2	2	-	34	26	20	1.378
3	3	8	-	2	2	-	40	23	18	1.331
4	3	8	-	2	2	-	40	23	20	1.404
5	3	8	-	2	2	-	40	26	18	1.462
6	3	8	-	2	2	-	40	26	20	1.528
7	3	8	-	2	5	-	34	23	18	1.232
8	3	8	-	2	5	-	34	26	18	1.337
9	3	8	-	2	5	-	34	26	20	1.383
10	3	8	-	2	5	-	40	23	18	1.335