Scientia Iranica, Vol. 13, No. 2, pp 152–158

SCIENTIA
IRANICA

Application of Rough Set Theory as a New Approach to Simplify Dams Location

M. Arabani* and M.A. Lashteh Nashaei[1]

Recently, advanced methods have been developed for selection of suitable sites for different types of dam. Apart from location, selection of the proper type of dam for a given site is of the greatest importance for engineers. Although these methods are being developed and some new approaches, like GIS techniques, are being used currently, all of these methods are mostly dependent on engineering decision making and a need for high costs. In this study, a new approach in the determination of dam locations is proposed. This method is based on the rough set mathematical theory that is a new approach presented by a well-known mathematician, Pawlak, in 1991. The information obtained from a practical case, the Karkhe project, which is one of the greatest dam projects throughout the middle-east, is used to present the procedure of this method. This dam was investigated in 1994 by famous consulting engineers and 30 places for the construction of this dam were considered. On the basis of the existing data for these places for siting of the Karkhe dam and based on the data required for the rough set method, finding a location for the most suitable site for this dam, using this new method, is undertaken and compared to the practical results obtained from site investigations. The results of this study are completely correlated with practical methods, which were performed by consulting engineers. The results also indicate that using this method for the project results in a total saving of 70% in costs.

INTRODUCTION

Feasibility studies of dams and reservoirs should always consider possible objections and take into account all relevant factors so that a proper effort can be made to mitigate any damages involved. Apart from the technical considerations, the environmental and ecological aspects of the project should be studied and provisions should be made to minimize any deleterious effects [1,2]. The objective of the feasibility stage is to obtain data for a cost estimate. The estimate should be sufficiently accurate to determine whether or not the project is economically justified. The accuracy of information required in the feasibility stage generally requires subsurface explorations to be performed. When thorough consideration has been given to each of the component studies for a particular project, they should be summarized, listing: (1) The favorable and (2) The unfavorable circumstances with regard to the project. Regarding all these considerations, in this paper, a new simplified method for the selection of suitable locations for dams is presented. This method deals with the application of the rough set theory in decision making and takes into account all relevant factors for the judgment procedure.

*. *Corresponding Author, Department of Civil Engineering, University of Guilan, I.R. Iran.*
1. *Department of Civil Engineering, University of Guilan, I.R. Iran.*

DEVELOPMENT AND APPLICATION

The rough set theory is a mathematical tool for dealing with vagueness or uncertainty. This theory was formulated by Zdzislaw Pawlak, a professor and research scientist at the Institute of Theoretical and Applied Informatics in the Polish Academy of Sciences in 1991 [3].

Rough set theory is a natural generalization of the "twin" theory (well known in interval mathematics). In both theories, set S is interesting:

- It can be the set of possible values of some quantity or:
- it can be a set of pixels that form an image.

In many real-life situations, there is only partial information about the set S:

- For some points, it is known for sure that s belongs to the set S;
- For some other points, it is known for sure that s does not belong to set S;
- For some points, it is not known whether or not s belongs to the (unknown) set S.

In this case, the only information known about set S is that it is "in between" the set, L, of all points that definitely belong to S and the set, U, of all points that may belong to S (i.e., about which it is not known for sure whether or not they belong to S): L is a subset of S and S is a subset of U. In other words, the available information about the (unknown) set, s, can be represented by a pair of sets, L and U, such that L is a subset of U [4].

When both lower and upper approximation sets, L and U, are intervals, one gets a twin. In knowledge representation, it is natural to consider more general sets defined by properties. Namely, if the only information that one has about the elements, s, consist of the values of n basic properties $P_1(s), \cdots, P_n(s)$, then, the approximation requires definition. In mathematical terms, the set algebra generated by the sets $S_i = \{s|p_i(s)\}$ (i.e., the smallest class that contains all these sets and which is closed under union, intersection and complement) is considered and pairs, L and U, of elements from this algebra are taken. Such a pair is called a rough set [3].

Rough set theory has the attention of researchers and theoreticians worldwide and has been successfully applied in fields ranging from medicine to finance [5]. It is to be noted that rough set theory is not, basically, a MCDM method and is different in concept. This method uses the definition of mathematical sets and subsets in analyzing procedures.

UTILIZATION OF ROUGH SET THEORY IN DECISION MAKING FOR DAM SITE LOCATION

Firstly, to be able to make use of the rough-set theory to detect effective parameters in relation to dam location, the germane data and information have to be gathered and provided. In this study, the mentioned information has been prepared and collected from a practical case, the Karkhe project, which is one of the greatest dam projects in the middle-east. This dam is located in the southern part of Iran, which was investigated in 1994 by famous consulting engineers and 30 sites were considered for construction of this dam [6]. The above information was gleaned by experts in accordance with the diagnostic method prepared by twenty experts in dam construction. Removing indistinctive conditional attributes in the diagnostic method, the author constructed Table 1 with 12 conditional attributes, such as (a) Quantity of river discharge; (b) Quality of river discharge and (c) Sedimentation of river. Then, the collected data becomes categorized as a table in which each row reflects the specifications of a particular location. Each column of the mentioned table indicates one of the characteristics considered for the location and the last column shows the suitability of that location for dam construction.

Secondly, to utilize the rough-set theory and analyze the information placed in the mentioned table, it is required to classify the information. So, each conditional attribute is provided with 4 classes, which show high, medium, low and no severity and the decision parameter (attribute) is classified by three levels, which describe high, medium and low suitability conditions, H, M and L. Consequently, by defining the specified levels and allocating a digit code to each defined attribute in the rows of the table, the classification of all attributes has been undertaken.

Table 2 shows the class numbers of conditional attributes and danger levels for 30 locations diagnosed by example. For example, location S_1 is classified into class number 1 of conditional attribute (a) and class number 2 of conditional attribute (b) and its severity level is diagnosed as "H". In other words, this table shows the relation between the class numbers of the conditional attributes of each location and its decision level or decision attribute. Such relations and such a table are called "decision rules" and "a decision table", respectively [7].

Determination of Minimal Decision Algorithm

Primarily, it is necessary to check whether the decision levels are compatible with 12 conditional attributes in Table 2, which shows the summary of diagnostic results by experts. The decision rules of all the locations were examined to find non-deterministic rules; i.e., locations which were classified into one and the same class under every conditional attribute but were assigned different decision levels.

Non-deterministic rules were not found in Table 2 and, hence, the decision level proved subordinate to conditional attributes. If non-deterministic rules are found in such a decision table, it means that the number of conditional attributes in the decision table is not sufficient and new conditional attributes have to be added to existing ones. In the process of extracting a minimal decision algorithm, it is necessary, for the time being, to use trial and error to rectify non-deterministic rules, if any, and make a decision table free from contradictions.

Alternatively, the locations, L_3, L_{13}, L_6, L_{18}, L_{10}, L_{15}, L_{11} and L_{24} were governed by one and the same rule. In such a case, it suffices to remove one location and consider only the other. Accordingly, the locations

Table 1. Conditional attributes for decision levels of the selected dam locations.

Conditional Attributes	Classification of Individual Situations	Decision Levels
(a) Quantity of river discharge	1- Flood in all seasons	*H*
	2- Flood during winter	*M*
	3- Normal base-discharged river	*L*
	4- Drought during summer	*N*
(b) Quality of river discharge	1- High-quality water (drinkable)	*H*
	2- Needs only physical treatment	*M*
	3- Muddy and polluted river	*L*
	4- Needs both physical and chemical treatment	*N*
(c) Sedimentation of river	1- Very low sedimentation	*H*
	2- Remarkable sedimentation	*M*
	3- Very high sedimentation	*L*
(d) Topography	1- Narrow-rocky valley	*H*
	2- Narrow-soily valley	*M*
	3- Wide valley	*L*
	4- Wide-alluvial valley	*N*
(e) Geology and tectonic	1- Very fine, compacted, rocky site	*H*
	2- Sedimentary layers	*M*
	3- Possibility of layers movement in site and reservoir	*L*
	4- Low quality soil with harmful minerals	*N*
(f) Access to construction materials	1- Materials available near site	*H*
	2- Materials can be transported quite easily to site	*M*
	3- Materials far from site	*L*
(g) Bearing capacity of foundation and embankment	1- Rocky with high strength	*H*
	2- Rocky with cracks	*M*
	3- Quite good soil (noncohesive)	*L*
	4- Alluvial bed with low bearing capacity	*N*
(h) Ground flooding	1- Without any structure in the reservoir	*H*
	2- Existence of roads in the reservoir	*M*
	3- Existence of farmland houses in the reservoir	*L*
	4- Existence of roads, farms, houses and other facilities in the reservoir	*N*
(i) Access to dam	1- Existence of access road	*H*
	2- Quite easy access to site by minor roads	*M*
	3- Need to construct access road to site	*N*
(j) Environmental impact	1- No environmental impact	*H*
	2- Creating some solvable environmental problems	*M*
	3- Severe environmental impact	*N*
(k) Location of hydraulic structures	1- Suitable location for all structures	*H*
	2- Rocky site with no suitable place for powerplant	*M*
	3- Costly hydraulic structures	*L*
	4- No place for construction of spillways and other hydraulic structures	*N*
(l) Economical considerations	1- Very economical site with benefit-damage balance	*H*
	2- Partly economical site with costly maintenance	*L*
	3- Non-economical site	*N*

Table 2. Observation data for diagnosis of dam location decision levels.

Locations	Conditional Attributes												Decision Levels
	a	b	c	d	e	f	g	h	i	j	k	l	
L_1	1	2	1	3	4	2	2	1	1	1	1	3	H
L_2	1	3	1	3	4	3	3	2	1	1	2	3	L
L_3	2	2	1	3	4	3	3	2	1	1	3	3	L
L_4	2	3	2	3	4	3	3	2	2	1	2	3	L
L_5	1	2	1	2	2	1	2	1	1	1	2	2	H
L_6	1	3	1	3	3	2	3	2	1	2	2	2	M
L_7	2	3	1	3	3	3	2	1	1	1	2	2	L
L_8	1	2	1	2	2	1	2	1	1	1	1	2	H
L_9	2	2	1	3	3	2	2	1	2	1	2	1	M
L_{10}	1	2	1	2	2	1	2	1	1	1	1	1	H
L_{11}	1	2	1	3	3	3	2	2	2	1	3	1	L
L_{12}	2	3	2	3	3	2	3	2	1	2	2	2	M
L_{13}	2	2	1	3	4	3	3	2	1	1	3	3	L
L_{14}	2	2	2	3	2	1	2	1	1	1	2	2	M
L_{15}	1	2	1	2	2	1	2	1	1	1	1	1	H
L_{16}	1	3	2	1	2	2	1	2	3	1	1	2	M
L_{17}	2	2	1	2	1	2	1	3	2	2	1	3	M
L_{18}	1	3	1	3	3	2	3	2	1	2	2	2	M
L_{19}	1	3	2	2	1	1	2	1	1	1	2	1	H
L_{20}	1	1	2	2	2	1	3	1	2	2	1	1	M
L_{21}	2	2	1	2	2	1	3	3	3	1	2	2	L
L_{22}	3	1	3	2	2	1	3	1	1	2	2	2	L
L_{23}	1	2	1	1	1	2	2	1	2	1	2	1	H
L_{24}	1	2	1	3	3	3	2	2	2	1	3	1	L
L_{25}	3	2	2	2	1	2	2	3	2	2	3	1	L
L_{26}	2	2	1	1	2	1	2	2	2	2	1	1	M
L_{27}	2	1	2	2	1	1	2	1	1	1	2	2	M
L_{28}	1	1	1	3	2	1	3	1	2	2	1	2	L
L_{29}	1	2	1	3	4	2	2	1	1	1	1	1	H
L_{30}	2	2	1	1	2	1	2	1	2	2	2	2	M

L_{13}, L_{15}, L_{18} and L_{24}, which are marked with "#", were removed from Table 2 to obtain a new decision table.

In order to find the insignificant conditional attributes in the diagnoses, a number of conditional attributes should be removed each time and the decision table should be checked to make sure any contradiction has not occurred. For instance, if one removes the conditional attributes (b), (f), (g), (h) and (k), the decision rules of locations S_1 and S_2 will be contradictory to each other, which means that the decision level of locations S_1 and S_2 is subordinate to one of the conditional attributes (b), (f), (g), (h) and (k). Thus, one cannot remove these conditional attributes simultaneously. Each combination was removed from Table 2, minus the locations L_{13}, L_{15}, L_{18} and L_{24}, then, it was checked whether any contradiction occurred among the decision rules.

Eight combinations of conditional attributes, cases 1 to 8, have been shown in Table 3. All of the combinations consist of the minimum number of conditional attributes, but, are still able to diagnose the problem. For illustrating the procedure for extracting the minimal decision algorithm, the method used for case 1 is presented.

In case 1 of Table 3, the conditional attributes, other than those of (a), (b), (d) and (f) of case 1, were removed from Table 2, minus the locations L_{13}, L_{15},

Table 3. Combinations of conditional attributes.

Number of Cases	Conditional Attributes			
1	(a)	(b)	(d)	(f)
2	(a)	(b)	(h)	(l)
3	(a)	(c)	(f)	(g)
4	(a)	(c)	(f)	(i)
5	(a)	(d)	(f)	(g)
6	(a)	(e)	(g)	(l)
7	(a)	(f)	(g)	(l)
8	(a)	(f)	(i)	(l)

L_{18} and L_{24}. Thus, a new table was obtained. In the new table, pairs of locations, L_1 and L_{29} and L_4 and L_7, etc. have been classified using one and the same rule for each. The locations of each pair were removed except one and the results are given in Table 4.

Reducing the Classes of the Conditional Attributes

Finally, the classes of the conditional attributes in Table 4 should be checked. All of the class numbers of each conditional attribute were removed one by one and checked for any contradiction. If there is any contradiction, it means that the removed class number is significant in the suitable location, otherwise, it is insignificant in the diagnosis. For instance, one can obtain a new decision table by removing class number 1 of the conditional attribute (a) of location L_1. In this table, location L_1 is assigned numbers 2, 3 and 2 in the columns of the conditional attributes (b), (d) and (f). The decision level is diagnosed as "H" and location L_9 is assigned to the same number in the columns of the same conditional attributes, while its decision level is diagnosed as "M". Thus, there is a contradiction between the decision rules of segments L_1 and L_9, which are marked with "#", meaning that the class number 1 of conditional attribute (a) of location L_1 is significant in diagnosing the decision level of location L_9 and, so, should not be removed.

As mentioned above, a new decision table is obtained that contains no contradiction and, although many of the class numbers have been removed, this table can still make the same diagnoses as Table 4. In this table, the two locations, L_6 and L_{12} and the same groups of locations, such as L_2, L_3 and L_4, etc. have been regarded as one and have the same rule. Therefore, the location of each pair or group was removed except one, in order to obtain a new decision table called Table 5. This table is the "minimal decision algorithm", in which there can be no

Table 4. Decision table based on conditional attribute in case 1.

Locations	Conditional Attributes				Decision Levels
	(a)	(b)	(d)	(f)	
L_1	1	2	3	2	H
L_2	1	3	3	3	L
L_3	2	2	3	3	L
L_4	2	3	3	3	L
L_5	1	2	2	1	H
L_6	1	3	3	2	M
L_9	2	2	3	2	M
L_{11}	1	2	3	3	L
L_{12}	2	3	3	2	M
L_{14}	2	2	3	1	M
L_{16}	1	3	1	2	M
L_{17}	2	2	2	2	M
L_{19}	1	3	2	1	H
L_{20}	1	1	2	1	M
L_{21}	2	2	2	1	L
L_{22}	3	1	2	1	L
L_{23}	1	2	1	2	H
L_{25}	3	2	2	2	L
L_{26}	2	2	1	1	M
L_{27}	2	1	2	1	M
L_{28}	1	1	3	1	L

Table 5. Minimal decision algorithm in case 1.

Locations	Conditional Attributes				Decision Levels
	(a)	(b)	(d)	(f)	
L_1	1	2	-	2	H
L_2	-	-	-	3	L
L_5	1	2	-	1	H
L_6	-	3	-	2	M
L_9	2	-	-	2	M
L_{14}	-	2	3	1	M
L_{19}	-	3	-	1	H
L_{20}	1	1	2	-	M
L_{21}	2	2	2	1	L
L_{22}	3	-	-	-	L
L_{23}	-	2	1	2	H
L_{26}	-	-	1	1	M
L_{27}	2	1	-	-	M
L_{28}	-	1	3	-	L

single conditional attribute or class removable without causing contradiction [8].

COMPARISON BETWEEN ROUGH SET AND PRACTICAL RESULTS

The frequency of appearance in Table 3, resulted from the rough set theory of each conditional attribute required by the minimal decision algorithm, is shown in Figure 1. The appearance frequency of the conditional attribute (a), which is required by the minimal algorithm of case 8, is eight. And, for the conditional attribute (e), which is only required by the minimal decision algorithm of case 6, its appearance frequency is one. By ranking conditional attributes in the descending order of appearance frequency, one can see that the most significant conditional attributes in evaluating the decision level of the location in this project are: (a) Quantity of river discharge; (f) Access to construction materials; (g) Ground flooding and (l) Economical considerations.

To evaluate this conclusion, the results obtained from the rough set theory are compared with the practical results determined by the consulting engineers of the Karkhe dam. This result is in accordance with the consulting engineers report in which the important factors of the Karkhe dam locations are listed [6]. As can be seen, the quantity of river discharge and access to construction materials are of the most importance in decision making for this particular case. Bearing in mind that the Karkhe dam is the second most highly-discharged river in Iran and that is an earth dam with a very high earthwork volume, the importance of easy access to clay and sand-gravel resources is clear. Also, geotechnical parameters and economical conditions are presumed to be key factors in the decision making and the report of the consulting engineers includes these two important factors in the selection of the Karkhe dam location.

As a general calculation, it can be revealed that by using the approach presented in this paper, the decision making factors were minimized into four important elements, which reduced the time and cost of the feasibility study and resulted in a saving of up to 70% in the project management. The reason for this conclusion is the ratio of four to twelve attributes, which can reduce the total calculation time to 30%.

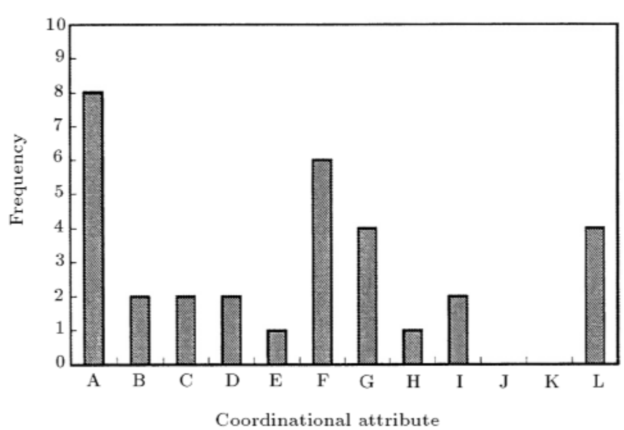

Figure 1. Frequency of conditional attributes.

The minimal decision algorithm can be described by rules such as "If conditional part (conditional attribute), THEN, conclusive part (decision level of distress or decision attribute). For example, one can describe the minimal decision algorithm of case 1 (Table 5) by 14 rules. The decision rules of locations L_1 and L_2 are represented by rules 1 and 2, respectively, and rule 14 expresses the decision rule of location L_{28}. The decision rule of L_2 in the tables indicates that if a dam were classified into class number 3 of conditional attributes (f), the decision level would be "L", or:

$$\text{if } (f) = 3 \text{ THEN decision level } = L.$$

When both of the conditions occur at the same time, and one of the conditions is that the location's classification is class number 3 of conditional attribute (b) and the other being when the classification of the location is class number 2 of conditional attribute (f), then, decision level is assigned as "M". The rule below describes location L_6 as:

$$\text{if (b)} = 3 \text{ and (f) } = 2 \text{ THEN decision level } = M.$$

In order to determine the conclusive decision levels using Expert Systems (ES), one needs to check if all the conditions in each rule were satisfied. For instance, for rule 12, one needs to check whether the location is classified into 1 of (d) and (f) or not. Altogether, one needs to check 33 conditions in 14 rules of case 1. Further, the decision tables of the minimal decision algorithm for cases 2 to 8, have also been extracted. Case 3 has the least number of rules and the most number of conditions.

CONCLUSIONS

In the current study, a rough set theory-based approach as a new method is presented and evaluated to determine dam locations using a complete practical example. Rough set theory is a powerful mathematical tool, that enables the users to decide and judge the minimal algorithms required for a particular problem, so that the accuracy of the selected algorithm is kept while the number of attributes is remarkably reduced. The major advantages of this method are: The simplicity of input data construction, the high speed and precision of the method during the recognition of the simplified algorithm and the avoidance of the formation of weighted matrixes for comparison of alternatives. Therefore, in this method, there is no need to compare parameters at the beginning and all relevant parameters are equally entered in the calculation procedure, which results in a remarkable saving in the time and cost of the calculation. The conclusions of this study are as below:

a) The advantages of the rough set theory in dam location present the high reliability of this method in the investigated example and can be introduced as a well-built useful mathematical based method;

b) By using the presented method, the cost of the process of dam location, in this particular case, can be reduced by up to 70% and can lessen the overall cost of the feasibility study;

c) The rough set theory is used for expressing a set of decision attributes by a set of paired conditional attributes. Thus, one can use this theory as a method of acquiring information from the diagnostic cases that exist in civil engineering problems.

REFERENCES

1. Ujorth, P. et al. "Operation monitoring and decommissioning of dams final version: November 2000", *Secretariat of the World Commission on Dams*, Vlaeberg, Cape Town 8018, South Africa (2000).

2. U.S. Army Corps of Engineers, "Design of small dams", (1998).

3. Pawlak, Z. "Rough sets: Theoretical aspects of reasoning about data", *Kluwer Academic*, Boston (1991).

4. Xury, Y., Zhihua, Z., Ning, L. and Shifu, C. "An approach for data filtering based on rough set theory", *Lecture Notes in Computer Science*, **2118**, Berlin: Springer-Verlag, pp 367-374 (2001).

5. Tsumoto, S. "Automated extraction of medical export systems rules from clinical databases, based on rough set theory", *Information Science*, **112**, pp 67-84 (1998).

6. Mahab Ghods (MG) Consulting Engineers, "Karkhe dam, the feasibility study report", Ministry of Power (2000).

7. Kryszkiewicz, M. "Properties of incomplete information systems in the framework of rough-sets", *Rough Set and Knowledge Discovery*, L. Polliowski and A. Skowron, Eds., Chapter 21, pp 422-450 (1998).

8. An, A.J. et al. "Discovering rules for water demand prediction: An enhanced rough set approach", *Eng. Applic. Artif. Intell.*, **9**(6), pp 645-653 (1996).