2. Cai Shaohuai (2002) Recent Development of Steel Tube Confined High Strength Concrete Structures in China. *Building science*, 18(4), p.p.1-7.
3. Yang Jie (2008) Research and development of steel tube concrete. *Engineering Structure*, 25, p.p.174-175.
4. Xu Chengxiang, Zhu Chang'an, Xu Shuzhen (2009) Experimental research on concrete-filled square thin-walled steel tubular stub columns with steel bar stiffeners. *Journal of Building Structures*, 2009(S2), p.p.231-236.
5. Han Linhai (2007) *Theory and practice of concrete filled steel tubular structure*, Beijing, Science Press.
6. Zhan Meisen (2007) *Research on Behaviour of Concrete-Filled Thin-Walled Steel Tube under Axial Compression*, Nanchang University
7. Zhang Yaochun, Wang Qiuping, Mao Xiaoyong, Cao Baozhu (2005) Research on Mechanics Behavior of Stub-column of Concrete-filled Thin-walled Steel Tube Under Axial Load. *Building Structure*, 35(1), p.p.22-27.
8. Gao Jianping, Wu Yonggen, Huo Jingsi (2012) Experimental study on the hysteretic behaviors of thin-walled corrugated concrete-filled steel tube column. *World Earthquake Engineering*, 28(3), p.p.34-34
9. Cao Baozhu, Zhang Yaochun (2004) Limit of width-thickness ratio of steel plates in concrete-filled square thin-walled steel tube columns. *Journal of Harbin Institute of Technology*, 36(12), p.p.1713-1716.
10. Chen Bo, Hao Zhe, Bai Huiwen, Zhao Yingna (2014) Corrosion Types under Different Conditions of Stainless Steels and the Present Research Status. *Total Corrosion Control*, 28(7), p.p.15-19.
11. A.V.Shanbhag, T.V.Venkatesha, B.M.Praveen, S.B. (2014) Abd Hamid,Inhibition Effects of Chloroquinolines on Corrosion Of Mild Steel in Hydrochloric Acid Solution. *Journal of Iron and Steel Research*, 21(8), p.p.804-808.

Decision Rule Extraction for Maritime Accidents in Inland Rivers Based on Rough Set

Shoucheng Deng

School of Logistics Engineering, Wuhan University of Technology, Wuhan 430063, China
School of Computer & Information, Three Gorges University, Yichang 443000, China

Qing Wu

School of Logistics Engineering, Wuhan University of Technology, Wuhan 430063, China

Xiumin Chu

Intelligent Transport System Research Center, Wuhan University of Technology, Wuhan 430063, China

Xianqiao Chen

School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China

Deshan Chen

Intelligent Transport System Research Center, Wuhan University of Technology, Wuhan 430063, China

Abstract

In view of data mining for the decision-making of the inland waterway transportation, we study a rough set based approach to extract the decision rules with historical maritime accidents data in the inland rivers. In order to solve the NP-hard problem of attribute reduction in Rough Set Theory, the genetic algorithm based attribute reduction is proposed and described in detailed steps. Noting that there are generally multiple sorts of decision-making in various types of historical accidents, the decision problem thus first needs to be specified and its relevant attributes need to be determined to construct the decision table. Then, the relative minimal reduct of the decision table can be calculated using the proposed heuristic reduction algorithm, so that the decision rules can be obtained. The effectiveness of the proposed method is demonstrated with the decision rule extraction on collision and grounding accidents in the key segment of the Yangtze River.

Key words: MARITIME ACCIDENT, ROUGH SET THEORY (RST), RULE EXTRACTION, GENETIC ALGORITHM, YANGTZE RIVER

1. Introduction

Inland waterway transportation has notable advantages in both aspects of economy and environmental protection. China has abundant resources of inland rivers for water transport. The Yangtze River and the Pearl River are two waterway arteries in China, the waterborne trade of them amounted to 2.06 and 0.7 billion metric tons in 2014, respectively. With the increasing volume of inland waterway transportation in recent years, safety assurance of maritime accidents in the inland rivers, in particular, emergency decision-making to them is facing a huge challenge.

The types of maritime incidents in the inland rivers include collision, stranding, fire, oil spill, explosion, etc. According to the statistics of Maritime Safety Administrations, there have been large volumes of maritime accidents in the inland rivers which caused losses of human life and property, or contaminated the environment. In this paper, we focus on how to conduct comprehensive analysis of the accidents data and discover the hidden knowledge in them so as to take measures to prevent the future accidents.

Rough Set Theory (RST) , first developed by Polish computer scientist Zdzislaw Pawlak [1] in the early 1980's, is a mathematical tool for quantitatively analyzing and processing uncertain, inconsistent and incomplete information. It has become another applicable method to handle uncertain problems besides Probability Theory, Fuzzy Sets and Evidence Theory. In RST, a rough set is basically an approximation of a crisp set in terms of two subsets, i.e., the lower and upper approximation of the original set. Its concepts are well suited for inducing the knowledge hidden in the information systems, which may be unraveled. Using the concepts of RST to represent and analyze knowledge from examples, it has the advantages that its algorithm of knowledge discovery depends on the data itself, and does not need other transcendental or extra information. Since being proposed, RST has become a fundamentally important approach in the field of artificial intelligence and cognitive science [2, 3]. Over the past years, a large number of extensional studies of rough set approach have been conducted in the literatures, such as the concept of decision-theoretic rough set [4], the dominance-based rough set approach in consideration of a multi-criteria classification problem [5], and the covering rough sets [6], etc. There also has been a rapid growth in its applications in research areas such as knowledge discovery, machine learning, data mining, feature selection, etc.

Rule Induction is one important research contents on RST. In the actual application, a large quantity of information and various uncertain factors are confronted. To apply RST can analyze the previous classification and division of empirical data, find out the valuable correlation among data and extract sets of decision rules. Grzymala-Busse [7] proposes three algorithms: LEM1, LEM2 and IRIM to induce rules from inconsistent data, learning from examples based on rough set. Ziarko & Shan [8] present a method for rule extraction by forming a decision matrix, after calculating the conjunction over all non-empty entries of the decision matrix, all disjunctions of minimal disjunctive form of the expression define the reducts, and the decision rules can be given out accordingly. Stefanowski & Tsoukias [9] study incomplete systems under the condition that all missing attribute values are lost to extend RST in inducing rules. Pawlak [10] gives the definition of the support, strength and certainty factors of the decision rule.

In this paper, we collect historical accidents data occurred in the Yangtze River in recent years, and present a rough set based approach for decision rule extraction to discover the implicit and tacit knowledge from the maritime accident database. The following sections are as follows: Section 2 introduces the basic concepts of RST; Section 3 proposes the rough set attribute reduction algorithm based on genetic algorithm, which is a key step for the rule extraction process; Section 4 gives the detailed process of inducing rules with historical maritime accident database. We examine our method on two types of maritime accidents of collision and stranding in the key segment of Yangtze River, and the case study demonstrates that the designed cognitive procedure would be appropriate to discover the tacit knowledge of decision rules in handling the historical maritime accidents data.

2. Relative concepts of RST

In this section, the basic concepts of RST are introduced to meet the requirement to understand our proposed application of RST into rule extraction from maritime database.

Definition 1 (Decision table). Let $T = (U, A)$ be a decision table, where:

$U = \{x_1, x_2, \dots, x_n\}$ is a non-empty finite set of cases.

$A = \{a_1, a_2, \dots, a_m\}$ is a non-empty finite set of attributes, and

$A = C \cup D$, C and D are the subsets of attributes, called the condition attribute set and decision attribute set respectively.

With the decision table, there is an associated equivalence relation, called P-indiscernibility relation, denoted by $IND(P)$ or U/P:

$$IND(P) = \{(x, y) \in U \times U \mid \forall a \in P, a(x) = a(y)\} \quad (1)$$

Where $P \subseteq A$, and (x, y) is the pair of objects from U. $a(x)$ denotes the value of attributes for object x. If $(x, y) \in IND(P)$, then objects x and y are indiscernible from each other in attributes from P.

The attribute values of a maritime accident contain mixed types of data such as symbolic, crisp numeric, interval numeric, and fuzzy linguistic data, etc.

When rough set theory is deployed to deal with the decision table, it requires the values from the processed decision table to be discrete. Therefore, the data must undergo the pre-processing of discretization that converts continuous attributes into discrete ones before using RST. For the discretization of the numerical attributes, a number of cut-points are determined to divide the attribute domain into consecutive subintervals.

Definition 2 (Lower and upper approximation). Based on P-indiscernibility relation, the target $X \subseteq U$ can be approximated using only the contained elements within P by constructing the P-lower and $\overline{P}$-upper approximations of X, denoted by $\underline{P}X$ and $\overline{P}X$ respectively:

$$\underline{P}X = \{x \mid [x]_P \subseteq X\} \tag{2}$$

$$\overline{P}X = \{x \mid [x]_P \cap X \neq \phi\} \tag{3}$$

where $[x]_p$ denotes the equivalence classes of P -indiscernibility relation.

Definition 3 (Positive region). Corresponding to the condition and decision equivalence attribute set C and D in the decision table T, the C-positive region of D is the union of all equivalence classes in $[x]_p$ which are subsets of the target set, denoted by $POS_C D$:

$$POS_C(D) = \bigcup_{X \in U/D} \underline{C}X \tag{4}$$

Definition 4 (certainty factor) Corresponding to $T = (U, C \cup D)$,let $P \subseteq C$,and P -indiscernibility relation relative to decision attribute set D is expressed as $IND(P) = \{Y_1, Y_2, \dots, Y_r\}$, we associate the certainty degree of C relative to D as:

$$k_P(D) = \sum_{i=1}^{r} \frac{card(\underline{P}Y_i)}{card(U)} \tag{5}$$

where $|X|$ denotes the cardinality of X. If it means that the decision attributes are determined by the condition attribute set P, and if $k_p(D) = 0$, be the opposite. If $0 < k_p(D) < 1$, it means that the decision attributes are dependent on the condition attributes set P in the degree of $k_p(D)$.

Definition 5 (Reduct and Core).

Let $R \subseteq C$, if $T' = \{U, R \cup D\}$ is independent, i.e., $POS_R(D) = POS_C(D)$, the attribute set of R is called a reduct of C, notated as $RED(C)$.

Additionally, the set of all indispensable attributes in C is called the core of C, notated as $CORE(C)$.

$$CORE(C) = \cap\, RED(C) \tag{6}$$

A reduct can fully characterize the knowledge of decision table which is not unique and may have alternative answers. The relative minimal reduct means the subset of attributes as small as possible. The core is the intersection of all the reducts, so the selected attributes in the core is common to all reducts.

3. Rough set attribute reduction based on genetic algorithm

Attribute Reduction is a main research content on RST, it has great significance on the rule extraction from the decision problems to seek for the minimal reduct. Because there always are multiple reducts for a decision table, the calculation of minimal attri- bute reduct is a typical NP-Hard issue [11]. And up to now, there is no general and accurate method for the solution. In order to find the relative minimal attribute reduction with minimal condition attributes, the heuristic algorithm becomes a kind of typical algorithm of to solve this problem.

Genetic algorithm was first proposed by professor Holland, with the mechanism of natural selection simulating biological evolution, and the optimal solution of random search simulating natural process. Genetic algorithm is a kind of heuristic search algorithm that is used in artificial intelligence of computer science to solve the optimization problem. Therefore, it shows relatively great advantage in solving NP-Hard problem. According to the multiple-attribute characteristic of the algorithm to extract the decision rules from the inland maritime accidents, the rough set attribute reduction algorithm based on genetic algorithm are proposed so as to obtain the minimal attribute reduction effectively.

The algorithm takes all the individuals in the group as the population. Besides, it uses selection operator, crossover operator and mutation operator to constitute the so-called genetic manipulation. The core contents of the algorithm contain five basic elements, i.e., chromosome coding, design of fitness function, setting of initial population, design of genetic manipulation and setting of controls parameter. The basic flow is as what is shown in Figure 1.

The description of rough set attribute reduction algorithm based on genetic algorithm is as follows:

Input: a decision table $T = (U, C \cup D)$, condition attribute set $C = \{c_1, c_2, \dots, c_n\}$ and decision attribute set $D = \{d_1, d_2, \dots, d_l\}$.

Output: the relative minimal reduct R for the decision table.

Step 1: use Formula (5) calculate the support degree $k_C(D)$of condition attribute C relative to decision attribute D.

Step 2: let $Core(C) = \emptyset$,get rid of an attribute $c \in C$ one by one. If $k_{C-c}(D) \neq k_C(D)$, then $Core(C) = Core(C) \cup \{c\}$,i.e., the core is $Core(C)$,if $k_{C-c}(D) = k_C(D)$,then $Core(C)$ is the minimal reduct; otherwise, conduct step 3;

Step 3: to randomly generate the initial population composed of the individuals represented by n (number of condition attributes) binary strings in the length of m.

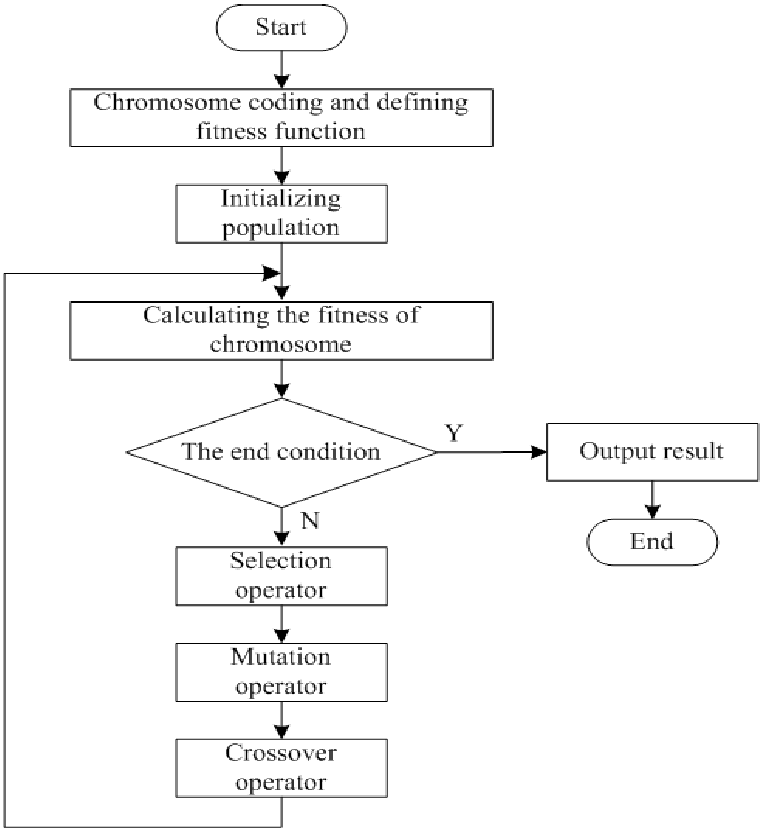

Figure 1. The iterative process of genetic algorithm

If the corresponding bit of condition attribute is the attribute in the middle of the core, the bit shall be 1; otherwise, 0 or 1 shall be selected randomly, and the fitness of each individual in the initial population shall be calculated by the following formula:

$$F(n) = 1 - \frac{L_n}{n} + k \tag{7}$$

In the formula, L_n is the number that the gene bit is 1 in n chromosomes; n is the number of condition attributes, that is, the length of chromosome; and $k_p(D)$ is the dependence degree of the decision attribute on the condition attribute contained in the chromosome.

Step 4: the roulette is used to choose individuals. A new generation of group is generated according to crossover probability p_c and mutation probability p_m. During variation, it shall always guarantee that the gene bit corresponding to the core attributes does not undergo variation;

Step 5: to calculate the relative fitness of each individual in the new generation of the group according to Formula (8);

$$CF(x_j) = F(x_j) / \sum_{j=1}^{m} F(x_j) \tag{8}$$

Step 6: the optimum maintaining tactics is adopted to copy the optimum individual to the next generation of the population;

Step 7: if the fitness value of the optimal individual in the continuous t generations does not rise any more, the calculation shall be ended; otherwise, Step 4 shall be turned to.

4. Process of decision rule extraction based on RST

The serious consequences caused by the maritime accidents and the negative impacts to the society are far more than the accident itself. How to analyze the accident objectively and scientifically, and find out the real reason of the accident after the accident occurs, so as to take measures to prevent future accidents is the problem that shall be solved currently. In the section, on the basis of collecting the maritime accidents in the key segments in Yangtze River during 2006 to 2011, making use of the database of maritime accidents, the process of decision rules extraction based on RST is applied, to explore knowledge concerning the multiple-factor rules of the maritime accidents in the aspects like accident time, accident ship, weather conditions like visibility, and accident type and so on, thus finally obtaining the decision rules.

To obtain rules, at first, it shall determine the decision problems, to extract the condition attributes and decision attributes concerning decision issues under the guidance of the experts in the domain field, the data about the relevant attributes are collected from the accident database. And then, the data are preprocessed, to determine the standard of data classification, and express the various attribute values in the process of discretization. From here, the decision table is formed, so that the attribute reduction algorithm based on genetic algorithm is adopted to calculate the reduct. At last, the decision rules are obtained according to the reduct, thus aiding to induce the knowledge of decision rules which are hidden in the accidents. The flow of extract rules extraction is shown in Figure 2.

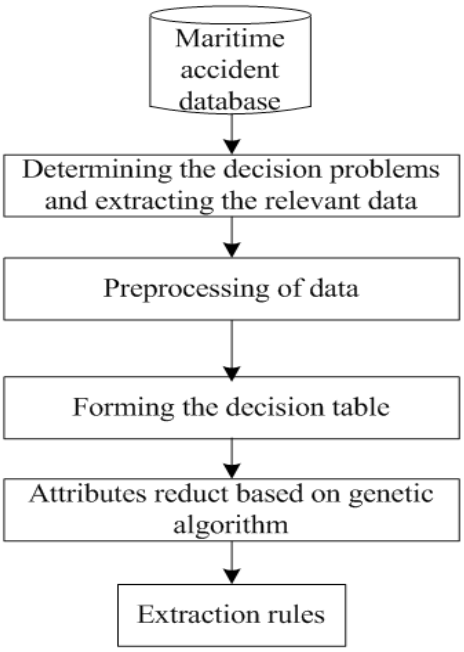

Figure 2. Process of the decision rule extraction

5. Instance analysis

In order to handle different types of data in the maritime accident database, it is necessary to carry out the discretization process before classification and approximation of the case set. The corresponding discrete variables of the attributes are defined as follows.

c_1 denotes wind speed: 1- below level 4, 2- level 4 or the above;

c_2 denotes rain: 1- no rain or light rain, 2- moderate or heavy rain, 3- rainstorm;

c_3 denotes wave: 1- no wave or light wave, 2- moderate to heavy wave;

c_4 denotes visibility: 1- above 500 meters, 2- below 500 meters;

c_5 denotes the water depth of the channel: 1- above 4 meters, 2- below 4 meters;

c_6 denotes water flow velocity: 1- no race, and 2- race or turbulent flow regime;

c_7 denotes the width of single-line navigation: 1- above 50 meters, 2- below 50 meters;

c_8 denotes whether the ship takes the wrong channel or not: 1- no, 2- yes;

c_9 denotes whether ship is overloaded or not: 1- no, 2- yes;

c_{10} denotes whether there is excessive draught or not: 1- no, 2- yes;

d denotes accident type: 1- collision, and 2- stranding.

The experiment is conducted towards the data in Table 1. Num signifies the number of accidents with the type d and its conditions of each line.

Table 1. The decision table of maritime accidents

No.	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	d	Num
1	1	1	1	2	1	1	1	1	1	1	1	52
2	1	1	1	2	1	1	2	1	1	1	1	3
3	1	1	1	1	1	1	1	1	2	1	1	9
4	1	1	1	1	1	1	2	1	2	1	1	1
5	1	2	1	1	1	1	1	1	1	2	1	8
6	2	1	2	1	1	1	1	1	1	1	1	4
7	1	3	1	1	1	2	2	1	1	1	1	1
8	1	1	1	1	2	1	2	2	1	1	1	10
9	1	1	1	1	1	1	1	1	1	2	2	30
10	1	1	1	1	1	1	1	1	2	1	2	5
11	1	1	1	2	1	1	1	1	1	1	2	15
12	2	2	2	1	2	1	1	1	1	1	2	1
13	1	2	1	1	2	2	1	1	1	1	2	2
14	1	1	1	1	1	1	1	2	1	1	2	32

Then, the genetic algorithm based attribute reduction is used to calculate the minimal reduct, and the parameter values are calculated as: m=15, p_c=0.8 and p_m=0.03. The optimal individual obtained by genetic algorithm is {1001010111}. Therefore, the obtained minimal attribute reduct is {wind speed, visibility, water flow velocity, taking wrong channel, overload, and excessive draught}. It can be conclu- ded that the five factors in the reduct set are relevant for leading to the collision and stranding accidents in the navigation segment.

On the basis of the obtained attribute reduct, the influencing factors of collision and stranding accidents are extracted in the form of decision rules, as shown in Table 2. Besides, according to the number of both types of accidents in the decision table, the confidences of each decision rule are also analyzed in the table.

Table 2. The decision rules obtained about collisions and stranding

No.	antecedent	Consequence1		Consequence2		consequence3	
		Decision attribute	Confidence	Decision attribute	Confidence	Decision attributes	Confidence
1	$a_1 = 2$	$d = 1$	4.7%	$d = 2$	1.8%	$(d = 1) \cup (d = 2)$	3.5%
2	$a_4 = 2$	$d = 1$	62.5%	$d = 2$	17.9%	$(d = 1) \cup (d = 2)$	45.1%
3	$a_6 = 2$	$d = 1$	1.1%	$d = 2$	3.6%	$(d = 1) \cup (d = 2)$	2.1%
4	$a_8 = 2$	$d = 1$	11.4%	$d = 2$	37.5%	$(d = 1) \cup (d = 2)$	21.5%
5	$a_9 = 2$	$d = 1$	11.4%	$d = 2$	5.4%	$(d = 1) \cup (d = 2)$	9%
6	$a_{10} = 2$	$d = 1$	9.1%	$d = 2$	33.9%	$(d = 1) \cup (d = 2)$	18.8%

The explanation of decision rules resulting to the reduct of condition attribute set are listed as below.

Rule 1: 4.7% of the collision accidents are caused by strong wind weather, and wind damage also leads to 1.8% stranding accidents. Although the total proportion of the two types of accidents aroused by wind damage is only 3.5%, it still shall not be ignored;

Rule 2: 62.5% collision accidents are caused by poor visibility, which also leads to 17.9% stranding accidents. The total proportion of the two types of accidents aroused by poor visibility is 45.1%. Poor visibility is the principal factor leading to the collision accidents, and also one of the main factors leading to the stranding accidents in the segment of Yangtze River.

Rule 3: The collision and stranding aroused by race or turbulent water flow account for 1.1% and 3.6% of the two types of accidents respectively. The total proportion of the two types of accidents aroused by race or turbulent water flow is only 2.1%.However, they should not be neglected.

Rule 4: Taking the wrong channel is the primary factor leading to the stranding of ships in this navigational segment, which accounts for 37.5% of the stranding accidents; and it also leads to 11.4% of the collision accidents. The total proportion of the two types of accidents aroused by taking the wrong channel is 21.5%.

Rule 5: ship overload mainly leads to collision accidents, which accounts for 11.4% of the collision accidents; and it also leads to 5.4% of stranding accidents.

Rule 6: Excessive draught of ships is one of the main factors leading to the stranding accidents of ships, which account for 33.9% of stranding accidents; besides, it easily leads to collision accidents. The total proportion of the two types of accidents aroused by excessive draught is 18.8%.

6. Conclusion

In recent years, Rough Set Theory has become the hotspot in the field of artificial intelligence, and it is widely applied in the domain of machine learning, knowledge acquirement and decision analysis. This article presented a preliminary study on knowledge discovery to induce the decision rules for inland maritime safety administration based on RST. The approach involves two sequential problems. First, the domain knowledge is required to specify the decision problem, clarifying the multi-attributes and the multiple or single object of the decision making problem, so as to extract the data about the specified decision problem from the database of historical maritime accidents. After the decision table is constructed, it needs to calculate the relative minimal reduct of the decision table. However, it is a NP-hard problem, and a genetic algorithm based attribute reduction was proposed to obtain the relative minimal reduct. The proposed algorithm combined the genetic algorithm and the relating concepts of classic RST, such as upper and lower approximation, reduct and core, certainty factor, etc., so as to extract decision rules with inland maritime accidents data. The instance analysis shows that the proposed method could extract valid decision rules from data, and the rules have probabilistic properties and could aid in preventing future accidents. It is expected that the induced rules would further enrich the knowledge base of intelligent decision-making support system for waterway transportation in the inland rivers.

Acknowledgements

This work was supported by the information technology project of Ministry of Transport, P.R.C (2013364548200).

References

1. Pawlak, Z. (1982) Rough sets. *International Journal of Computer & Information Sciences*, 11(5), p.p.341–356.
2. Pawlak, Z., & Skowron, A. (2007) Rough sets: Some extensions. *Information Sciences*, 177(1), p.p.28-40.
3. Pawlak, Z., & Skowron, A. (2007) Rudiments of rough sets. *Information Sciences*, 177(1), p.p.3-27.
4. Yao, Y. (2007) Decision-theoretic rough set models. *Proc. Conf. on Rough Set and Knowledge Technology*. Toronto, Canada, p.p.1-12.
5. Greco, S., Matarazzo, B., Slowinski, R. (2002) Rough approximation by dominance relations. *International Journal of Intelligent Systems*, 17(2), p.p.153-171.
6. Chen, D., & Wu, K. (2012) Properties of Degree Induced Covering Rough Set. *Lecture Notes in Electrical Engineering*, 135, p.p.393-397.
7. Grzymala-Busse, J. (1987) Learning from examples based on rough multisets. *Proc. Conf. on Methodologies for Intelligent Systems*. Charlotte, USA, p.p.325–332..
8. Ziarko, W., Shan, N. (1995) Discovering attribute relationships, dependencies and rules by using rough sets. Proc. Conf. on System Science. Hawaii, USA, p.p.293–299.
9. Stefanowski, J., & Tsoukias, A. (2001) Incomplete information tables and rough classification. *Computational Intelligence*, 17(3), 545-566.
10. Pawlak, Z. (2002) Rough set theory and its applications. *Journal of Telecommunications and Information Technology*, No.3, p.p.3:7-10.
11. Mafarja, M., & Abdullah, S. (2013) Investigating memetic algorithm in solving rough set attribute reduction. *International Journal of Computer Applications in Technology*, 48(3), 195-202.

Optimization of Land Utilization and Distribution Based on Multi-Objective Ant Colony Algorithm

Yue Wang, Yuyang He*, Xudi Chen

School of civil engineering and architecture, Southwest Petroleum University , Chengdu 610500, Sichuan, China
**Corresponding Email: 676012778@qq.com*