



A multivariable optical remote sensing image feature discretization method applied to marine vessel targets recognition

Mengxing Huang^{1,2}  · Qiong Chen^{1,2} · Hao Wang³

Received: 23 June 2018 / Revised: 7 April 2019 / Accepted: 21 June 2019

Published online: 28 August 2019

© The Author(s) 2019

Abstract

The effective extraction of continuous features in ocean optical remote sensing image is the key to achieve the automatic detection and identification for marine vessel targets. Since many of the existing data mining algorithms can only deal with discrete attributes, it is necessary to transform the continuous features into discrete ones for adapting to these intelligent algorithms. However, most of the current discretization methods do not consider the mutual exclusion within the attribute set when selecting breakpoints, and cannot guarantee that the indiscernible relationship of information system is not destroyed. Obviously, they are not suitable for processing ocean optical remote sensing data with multiple features. Aiming at this problem, a multivariable optical remote sensing image feature discretization method applied to marine vessel targets recognition is presented in this paper. Firstly, the information equivalent model of remote sensing image is established based on the theories of information entropy and rough set. Secondly, the change extent of indiscernible relationship in the model before and after discretization is evaluated. Thirdly, multiple scans are executed for each band until the termination condition is satisfied for generating the optimal number of intervals. Finally, we carry out the simulation analysis of the high-resolution remote sensing image data collected near the coast of South China Sea. In addition, we also compare the proposed method with the current mainstream discretization algorithms. Experiments validate that the proposed method has better comprehensive performance in terms of interval number, data consistency, running time, prediction accuracy and recognition rate.

Keywords Remote sensing image · Feature discretization · Vessel targets recognition · Information entropy · Rough set

✉ Mengxing Huang
huangmx09@163.com

¹ State Key Laboratory Marine Resource Utilization in South China Sea, Haikou 570228, China

² College of Information Science and Technology, Hainan University, Haikou 570228, China

³ Big Data Lab, Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

1 Introduction

Automatic detection and recognition of vessel targets is one of the most active research topics in the field of ocean remote sensing image analysis and processing. As its name implies, the aim of vessel detection and recognition is to extract, identify and locate ship targets in the remote sensing image without human intervention [18]. Among the data acquired on remote sensing satellites for vessel target surveillance, optical remote sensing image combined with the characteristics of all-day and all-weather SAR imaging including the phase information, has become the research hotspot of the current vessel target detection and recognition technology [7], due to their advantages of high spatial resolution, intuitive content, significant structure, etc.. To analyze and process the monitored data, the feature extraction of remote sensing images must be performed firstly. Since many existing data mining algorithms can only deal with discrete attributes, continuous features need to be transformed into discrete features to adapt to these intelligent algorithms for expanding the scope of application. On the other hand, the feature extraction of ship targets is mainly confronted with the problems of strong sea clutter interference in extreme sea conditions [3], numerous types of vessels, complex movement of vessels on the sea surface, less actual measured data of vessels and so on. In addition, the grayscale and texture features of remote sensing images are often indistinguishable from the port surface [10]. Compared with the offshore vessels on a simple sea background, the extraction of features is relatively more difficult. Therefore, reasonable discretization is very important in the process of feature extraction. It can not only reduce the space dimension of continuous features, eliminate data redundancy, reduce the complexity of program execution, but also reduce the loss of important information and ensure classification prediction accuracy, helps to improve the efficiency of subsequent intelligent detection and recognition algorithms.

The essence of discretization is simply to decide how many segmentation points to exploit and determine the segmentation point location. There are many methods about discretization. According to whether the data contains category information, they can be classified into supervised discretization [49] and unsupervised discretization [4]. Supervised discretization needs to consider category information, such as 1R [1], ChiMerge [34], etc., however, unsupervised discretization does not require any category information, such as Equal-Width [30], Equal-Frequency [11], and so on. Although discretization, as a data reduction technique in the data preprocessing stage, has received extensive attention and research in recent years, and has achieved fruitful research results [13, 36], however, most discretization algorithms still have relatively few applications in the analysis and processing of ocean optical remote sensing images, and all of them have certain defects, mainly in the following aspects: (1) many redundant breakpoints in properties and the lack of necessary breakpoints, which make the learning inaccurate; (2) get a very large number of intervals while avoiding information loss, and thus the overfitting phenomenon occurred; (3) exponential growth of the program complexity, unable to meet the real-time dynamic target recognition processing; (4) the choice of breakpoint does not consider the breakpoint mutual exclusion among the attributes and within the attribute, resulting in the destruction of the decision system compatibility; (5) the reason that difficult to obtain the prior knowledge about the sea or the large changes in the marine environment for ages leads to the fact that prior knowledge is no longer applicable, which makes the accuracy of the algorithm decreased. Based on the above analysis, these algorithms are not suitable for processing multi-feature optical remote sensing data in complex marine environment obviously.

Regarding the issues above, in order to get the optimal set of discrete breakpoints from the image and to quickly and accurately separate the vessel targets in the image, we propose a new method called MFD-mvtR (Multivariable optical remote sensing image Feature Discretization applied to marine vessel targets Recognition) in this paper. The basic idea is as follows: (1) tag the target area with significant visual features in the image, in the case of port images, the grayscale value of the boundary area of the port needs to be converted before the tagging to improve the range of gray areas of interest; (2) use the labeled area as a training sample to establish the image information decision table; (3) adopt the Top-Down discretization method for each band in the image to calculate the information entropy of all the intervals in the current band [22], then select the interval with the highest entropy value for splitting; (4) discretize the original decision table by the obtained candidate breakpoints, and introduce the equivalent model of rough set [27] to compare the upper and lower approximation sets of the original decision table with that of the new decision table to get the extent of change in the indiscernible relationship of the image information table; (5) adjust the algorithm parameters and the segmentation threshold according to the extent of the change in the indiscernible relationship, then rescan each band until termination condition are met to obtain the optimal discretization result.

In the original entropy algorithm [35], the number of segments is generally determined by the user-defined splitting number or the given minimum entropy threshold, while in our algorithm, besides the above conditions, the number of segments is controlled by calculating the number of differences about the upper and lower approximate sets between before and after discretization. In the literatures [9, 20, 51], when using the rough set to measure the system compatibility, Eq. 1 is usually used to calculate the dependence among the knowledges. When U is the set of objects, Q and R are knowledges about U , $POS_Q(R)$ is the positive domain of knowledge R under the representation of knowledge Q , and $card(\bullet)$ is the cardinality of the set. That is, the number of elements contained in the set.

$$\gamma_Q(R) = \frac{card(POS_Q(R))}{card(U)} \quad (1)$$

However, in practical applications, γ simply reflects the number of missing elements, while the upper and lower approximate sets describe the entire equivalence class. Its changes are directly related to the category information in the remote sensing image. Therefore, it would be more appropriate than γ to measure system compatibility.

Finally, the high-resolution remote sensing image data collected from the port area of the South China Sea is simulated and analyzed, and the proposed method is compared with EDiRa [35], ChiMerge [34], 1R [1], NCAIC [47], FUDC [50], Cramer's V-Test [43], Chi2 [31], these seven state-of-the-art discretization methods. Experimental results show that the proposed method not only has better comprehensive performances in terms of interval number, consistency, and prediction accuracy, but also achieves higher detection rate and lower false alarm rate in classifier of ship identification [38]. It validates the effectiveness of the proposed method in the application of marine ship target detection and identification. Therefore, our method is more suitable for the discretization of optical remote sensing image features for target detection and identification of marine vessels.

This paper is composed of four section. The remaining sections are organized as follow. Section 2 describes the problem model and basic concepts for the proposed work. Section 3

introduces the proposed algorithm model. The experimental results and discussion are presented in Section 4. Section 5 concludes this paper.

2 Problem model

2.1 Remote sensing image feature discretization

In simple terms, remote sensing image feature discretization is to adopt a specific method to divide a continuous feature interval on the image into a limited number of cells, then to associate these cells with a set of discrete values. The discretization of continuous features (also called continuous attributes) is an important preprocessing step for data mining and machine learning, and is directly related to the effect of mining or learning [6, 19].

The continuous features of optical remote sensing images are generally represented by digital number (abbreviated as DN values). According to different levels of quantification on different types of sensors, the value range of features in each band is not the same. Some use 8-bit quantization, then the DN value range is 0–255, and some use 16-bit quantization, such as high-resolution worldview-2 satellite [46], the value range is larger, reaching 0–65,535. On the other hand, there are many bands in multispectral remote sensing image, especially hyperspectral remote sensing images, the number of bands is as high as tens or even hundreds. As a result, in the features processing of optical remote sensing images, a TB-level data volume will be generated, which causes considerable difficulties for most of the knowledge extraction, data mining, classification and target recognition algorithms [5]. Therefore, it is very necessary to properly discretize the band pixel values in optical remote sensing images. It can convert quantitative data into qualitative data to obtain remote sensing feature partitions that do not overlap each other, also greatly reduce the amount of data to be processed, and optimize the data set [32].

Besides the above mentioned issues of large-scale data, the problem of data similarity is also very important. In the application of marine vessel target identification, due to the polymorphism of the port and the complexity of the background, the grayscale and texture features of the docking vessel are very similar to the ports, and they are difficult to distinguish in terms of the tonality of the image, which has caused great difficulties for the identification of boats in ports. However, through observing the pixel values of each band of the high-resolution optical remote sensing image, it is found that the pixel values of boats and ports are similar on some bands while significant differences on other bands, as shown in Fig. 1.

The above is a partial sample of the boat and port targets from the GF-2 satellite image with four bands. We can see that the DN values of boats and ports in band 1 are very close while significant differences in band 2, band3 and band 4. In general, the resulting knowledge granularity tends to be fine if the equivalence classes are divided in the bands with close DN values among different categories. On the contrary, the resulting knowledge granularity will appear coarse if the equivalence classes are divided in the bands with significant differences in DN values among different categories. When these bands are mixed together to divide equivalence classes, the bands with large differences will be affected by the bands with small differences, and the overall knowledge

: ROI name: boat						: ROI name: port					
: ROI rgb value: {255 0 0}						: ROI rgb value: {128 0 0}					
: ROI npts: 10						: ROI npts: 10					
File X	File Y	B1	B2	B3	B4	File X	File Y	B1	B2	B3	B4
4354	2343	0.316169	0.31157	0.187685	0.201234	3922	3828	0.362622	0.412668	0.283062	0.303552
4356	2343	0.342704	0.286512	0.173978	0.201234	3922	3829	0.3597	0.409943	0.280049	0.303552
4357	2343	0.316169	0.310181	0.183118	0.187844	3921	3829	0.364374	0.415392	0.282309	0.304661
4353	2343	0.309059	0.308097	0.187685	0.203465	3921	3830	0.352091	0.400393	0.270252	0.298006
4358	2344	0.332703	0.328209	0.19225	0.191193	3922	3830	0.353263	0.400393	0.277036	0.299116
4358	2345	0.337413	0.340654	0.102132	0.209041	3930	3830	0.318535	0.451692	0.230199	0.252468
4360	2345	0.389432	0.264839	0.156432	0.159915	3931	3830	0.327987	0.458579	0.237013	0.261363
4357	2345	0.310245	0.283023	0.17474	0.193425	3932	3830	0.340354	0.383296	0.253648	0.270253
4355	2346	0.300151	0.29	0.17474	0.185612	3921	3831	0.3597	0.405852	0.281555	0.299116
4357	2346	0.367875	0.362706	0.199093	0.21127	3922	3831	0.35619	0.401758	0.280049	0.298006

(a) DN values of boat

(b) DN values of port

Fig. 1 Comparison of DN values of boat and port in each band

granularity will be skewed toward the bands with small differences, which leads to the generation of excessive intervals to fail to achieve the ideal discretization scheme. Therefore, in addition to converting DN values at the port junction, we also need to group the bands for preventing the bands with large DN value differences of targets from being interfered by the bands with small DN value differences of targets in the process of discretizing the features of harbor images. For all remote sensing characteristics listed above, this paper establishes a basic framework of optical remote sensing image feature discretization for marine vessels recognition, as shown in Fig. 2.

First of all, the remote sensing image features are grouped according to the similarity of the pixel values of boat and port, and the new features set generated is sorted according to a certain specified rule, such as insertion sort, bubble sort, selection sort, quick sort, heap sort, shell sort, etc. Then, initially determine the dividing points of the continuous features, that is, the selection of initial breakpoints. The next step is to split or merge breakpoints according to the discretization algorithm. Finally, the discretization

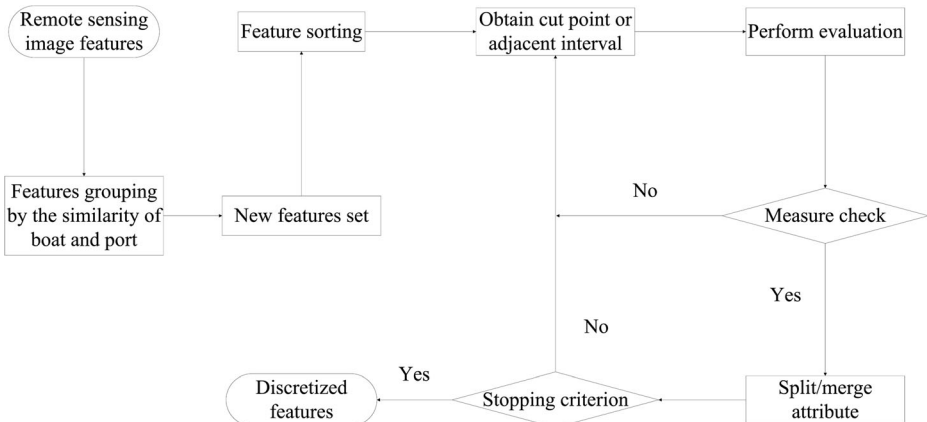


Fig. 2 Feature discretization process of remote sensing image

result is evaluated. If the criterion is satisfied, the whole discretization process is terminated, otherwise, returns to the previous step.

2.2 Remote sensing image feature model based on rough set

Rough set theory is an important mathematical tool for handling uncertain data [21]. In rough set theory, knowledge is regarded as the division of the universal, that is, knowledge is considered to be granular, and the uncertainty is caused by the large granularity in the knowledge. Different from the DS evidence theory [39] and the fuzzy set theory [14, 24], the membership function value of the object in the rough set theory depends on the knowledge base. It can be directly obtained from the required data without any prior knowledge or additional information. So, when the prior knowledge of the ocean is not easy to obtain, it is much more objective to use rough set to reflect the uncertainty of marine knowledge [26].

In rough set, data tables are called information systems. It can be described as a 4-tuple $S = (U, A, V, f)$, where U is a non-empty finite object set, A is a non-empty finite attribute set, $V = \cup(V_a)$ is a set of attribute values, and V_a is a value domain of attribute a , $f: U \times A \rightarrow V$ is a mapping function that represents the mapping from each object to an attribute value. If one of the attribute set is considered as a decision attribute, the above-defined information system S is called a decision table, where $A = C \cup D$ contains condition attribute set C and decision attribute set D .

Since optical remote sensing images generally contain multiple bands, i.e. multiple feature variables. If bands are discretized independently, the result will largely destroy the compatibility of the original system, thus affecting the subsequent classification accuracy and target recognition rate. Therefore, this paper establishes a multivariate remote sensing image feature model based on the rough set theory in the analysis and processing of remote sensing images. Where U denotes the collection of image pixels, the attributes in condition attribute set C represent bands, D contains only one decision attribute that corresponds to the land cover class in the remote sensing image, V_a represents the value domain of the a th band. The model is represented by the following matrix.

$$DS = \begin{bmatrix} u_1 & c_{11} & c_{12} & \cdot & \cdot & \cdot & c_{1m} & d_1 \\ u_2 & c_{21} & c_{22} & \cdot & \cdot & \cdot & c_{2m} & d_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ u_n & c_{n1} & c_{n2} & \cdot & \cdot & \cdot & c_{nm} & d_n \end{bmatrix} \quad (2)$$

Each row represents a sample item, training sample set $U = \{u_1, u_2, \dots, u_n\}$, vector $C = \{c_1, c_2, \dots, c_m\}$ indicates the DN values of the sample in m bands. The last column is the decision attribute column D , which identifies the category information of the sample. Each item consists of a sample number, band attributes, and a class attribute. The value range of band is $0 \leq c_{ij} \leq 1$, where c_{ij} is the DN value of the i th sample in the j th band. The value of the decision attribute is represented by a natural number. Its value range is determined according to the number of the given number of the categories. For example, if the number of defined categories is 5, the value range is $D = \{1, 2, 3, 4, 5\}$.

2.3 Information entropy measure of feature interval

Information entropy is a well-known mathematical theory proposed by Shannon, the father of information theory, for solving the quantitative measurement of information in the communication field [37]. Catlette, Fayyad, and Irani introduced information entropy into the discretization algorithm [2, 8]. According to the discussion of Fayyad and Irani, the formulas of information entropy and break point information entropy are given respectively.

$$E(S) = - \sum_{i=1}^k P(C_i, S) \log(P(C_i, S)) \quad (3)$$

$$E(A, T, S) = \frac{|S_1|}{|S|} Ent(S_1) + \frac{|S_2|}{|S|} Ent(S_2) \quad (4)$$

Where S is a set of objects, k is the number of categories, C_i represents the number of instances whose category is i in the set of objects S , A, T represents the breakpoint T on the attribute A , S_1 and S_2 represent the two objects sets of interval divided by breakpoint T respectively, $|S|$ denotes the cardinality of the set S .

The information entropy is a good measure for evaluating the divided feature intervals. It can reflect the stability of the frequency of all classes within the interval [40], thus ensuring the validity of the interval division. In literature [42], a semi-supervised classification framework of hyperspectral images based on the fusion evidence entropy is proposed and implemented by estimating the fusion evidence entropy of unlabeled samples using the minimum trust evaluation and maximum uncertainty, which makes it possible to achieve better classification charts with few labeled samples. Therefore, this paper applies information entropy to the evaluation of the feature interval division of optical remote sensing images. Where S denotes a set of image pixels, k denotes the number of land cover categories, C_i denotes the number of instances of the category i in the pixel set S , A, T represent the break point T in the band A , S_1 and S_2 represent the two pixel sets of interval divided by the break point T in the band A respectively, $|S|$ represents the cardinality of the set S , that is, the total number of pixels included in S [44].

3 Multi-variable optical remote sensing image feature discretization algorithm based on information entropy

The essence of discretization is to decide how many segmentation points to exploit and determine the segmentation point location, and then divide the subintervals or merge breakpoints according to certain criteria. The feature discretization method of remote sensing image based on information entropy proposed in this paper, is a multivariate

supervised algorithm that adopts the top-down [17, 36] strategy. The method is to find the one with the largest entropy among the subintervals each time, and then to gain the optimal number of intervals based on the indiscernible relationship.

3.1 Interval entropy table

In order to quickly find the subinterval with the largest entropy, a table needs to be established to record the entropy of all current intervals, i.e., the interval entropy table, abbreviated as IET. IET contains a total of 3 columns, the first column records the lower bound value of the corresponding interval whose upper bound value is recorded in the second column, and the third column records the corresponding entropy value obtained through a series of calculations, as shown in Table 1.

Each row in the table corresponds to a subinterval, and all of the subintervals are arranged in ascending order according to entropy. The method is to search for the separable interval with the largest entropy from the last item each time. Separable intervals contain at least two breakpoints (i.e., the lower bound of the interval is not equal to the upper bound of the interval), and the entropy is greater than the given threshold. At the beginning, IET contains only one row, that is, the entire continuous feature interval. As the algorithm runs, it starts to split. In the end, IET is implemented for saving all intervals by updating the minimum and maximum of the two operated intervals after adding a new row at the current split interval.

3.2 Calculating the number of differences in approximate sets

In order to calculate the number of differences between before and after discretization, the concepts of indiscernible relationship, lower approximation set and upper approximation set need to be introduced.

3.2.1 Indiscernible relationship

Given a decision table $S = (U, R, V, f)$, where U is a finite set of objects, R is a set of attributes, including a set of conditional attributes C and a set of decision attributes D . For each attribute subset $A \subseteq R$, the indiscernible relationship $IND(A)$ is defined in Eq. 5.

$$IND(A) = \{ \langle x, y \rangle \mid \langle x, y \rangle \in U^2, \forall a \in A (a(x) = a(y)) \} \quad (5)$$

Table 1 IET structure

Low Bound	Upper Bound	Entropy Value
L1	U1	V1
L2	U2	V2
⋮	⋮	⋮
⋮	⋮	⋮
Ln	Un	Vn

The equivalence class about attribute subset A in the universal U is also defined.

$$U|IND(A) = \{X|X \subseteq U \wedge (\forall x \in X \forall y \in X \Rightarrow \forall a \in A (a(x) = a(y)))\} \tag{6}$$

3.2.2 Lower approximate set and upper approximate set

According to the above decision table S , for each subset $X \subseteq U$ and the equivalence classes of the attribute subset A in the universal U , the lower and upper approximate sets of X are respectively defined in Eq. 7 and Eq. 8.

$$A_-(X) = \cup\{Y|Y \in U|IND(A) \wedge Y \subseteq X\} \tag{7}$$

$$A^-(X) = \cup\{Y|Y \in U|IND(A) \wedge Y \cap X \neq \emptyset\} \tag{8}$$

In order to elaborate on the calculation process of the lower and upper approximate sets differences between before and after discretization in the next section, we suppose that $A = C$, $X \in U | IND(d)$, and d is one of the decision attributes in set D . From the above definition, the lower approximate set $C_-(d_x)$ and the upper approximate set $C^-(d_x)$ corresponding to each decision attribute value can be calculated.

3.2.3 Differences between before and after discretization

According to the above definition, the number of differences $N_d = N_l + N_u$ between before and after discretization about the lower and upper approximate sets can be obtained, where N_l is the number of the lower approximate sets differences while N_u is the number of the upper approximate sets differences. N_d , N_l and N_u are initialized to 0 respectively. The calculation steps of N_d are as follows.

- Step 1: Select element $d_i(i = 1, 2, \dots, n)$ from decision attribute d of the original table, where n is the number of different values of the decision attribute d in the universal U , namely the number of categories;
- Step 2: Calculate the lower approximate set $C_-(d_i)$ and the upper approximate set $C^-(d_i)$ of d_i . If there are still elements in decision attribute d that have not been calculated, return to Step 1, otherwise, continue the next step;
- Step 3: Discretize the original decision table using the finally generated IET to get the new decision table $S^E = (U, R, V^E, f^E)$;
- Step 4: Select the element $d_i(i = 1, 2, \dots, n)$ from the decision attribute d in the new table, then calculate the lower approximate set $C_-(d_i)'$ and the upper approximate set $C^-(d_i)'$ of d_i ;
- Step 5: Determine whether the upper and lower approximate sets are equal before and after discretization, respectively, in case $C_-(d_i)' \neq C_-(d_i)$, then $N_l = N_l + 1$, in case $C^-(d_i)' \neq C^-(d_i)$, then $N_u = N_u + 1$; if there are still elements in the decision attribute d that have not been calculated, return to Step 4, otherwise, $N_d = N_l + N_u$, and the program ends.

3.3 Algorithm flow

Algorithm 1: MFD-mvtR Pseudo Code

Input: Remote sensing image features

Output: Discretized features

begin

```

New features set ← Features are grouped according to the similarity of the DN values of boat and port;
for each group i in new features set do
  while Stopping criteria are not met do
    for each feature j in group i do
      Create an IET table;
      Initialize IET by computing entropy values of all interval using Eq. (3);
      N ← Size of IET;
      k = N;
      while k > 0 do
        Arrange IET in ascending order according to entropy value;
        VE ← Entropy value of interval k;
        if VE < threshold do
          break;
        else if the number of breakpoints in interval k is larger than 1 do
          Split interval k using Eq. (4);
          k = N+1;
        else do
          k = k-1;
        end
      end
    end
    Modify the threshold;
    Stopping criteria ← Calculate the differences of the upper and lower approximation sets using Eq. (5) ~ Eq. (8);
  end
end
Return the discretized features from all groups;
end

```

A basic flow of MFD-mvtR algorithm is represented in Algorithm 1. At the beginning, the original decision table is input to the program execution and bands are grouped to generate the new features set according to the similarity of boat and port. Discretization is performed in order from the first attribute in the condition attribute set to establish the IET. Then, the separable interval with the largest entropy value is found to split from IET in each loop until all the attributes have finished the discretization. Finally, the number of differences between before and after discretization about the lower and upper approximate sets is calculated. If the specified deviation are not satisfied, then the splitting terminated conditions including the threshold of entropy and the number of iterations are modified, and the new features set will be re-discretized.

4 Experiments and analysis

4.1 Data source

The experimental data used in this paper comes from a GF-2 satellite data in the offshore port area, China, on October 7, 2015, as shown in Fig. 3. The multispectral image of this GF-2 satellite data contains four bands. The objects in this image are divided into six categories: boat, port, building, bare land shoal, water body and vegetation.

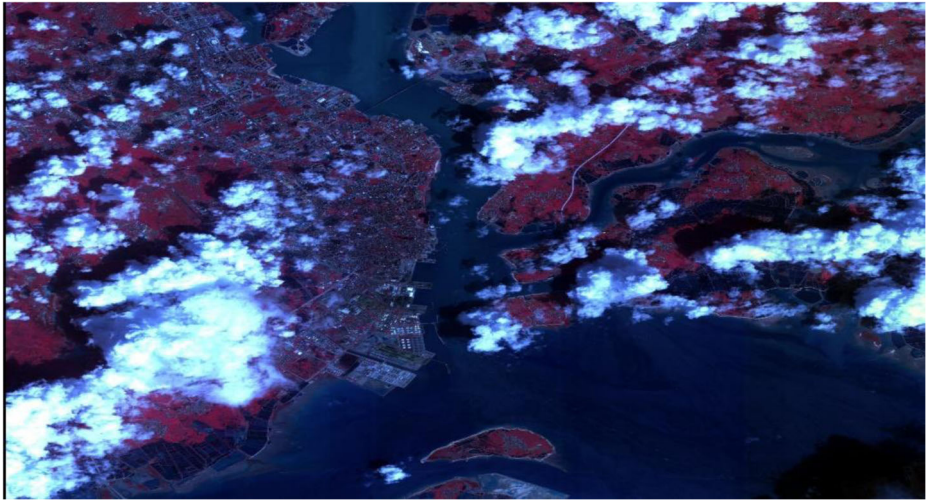


Fig. 3 Area used for study

4.2 Experimental environment

In order to verify the effectiveness of the proposed method, all four algorithms were executed on a computer with Intel(R) Core(TM) i5-5200 U CPU@2.20GHz processor and 12G RAM hardware. Visualization, programming, simulation, testing and numerical calculation processing of this experiment are implemented in MATLAB (R2016a version) environment. Radiometric calibration of images, atmospheric correction, and comparison of results before and after discretization are performed under ENVI 5.3 environment.

4.3 Evaluation of discretization quality

Firstly, several regions covering six major categories are randomly selected from the image and integrated as training samples to be discretized, containing a total of 2607 pixels, among which 676 are boats, 742 are ports, 143 are buildings, 116 are bare land shoals, 807 are water bodies, 123 are vegetation. Then, after the pixels are sorted, and eliminates the duplicates by value within the band, the number of initial breakpoints for the four bands is obtained, which is 502, 493, 358, 359, respectively. Therefore, the training sample has a total of 1712 breakpoints at the beginning. The quality of the discretization scheme mainly depends on the number of the obtained intervals and the data inconsistencies in the new information table. The number of data inconsistencies is expressed by the following mathematical formula.

$$Inconsistencies = \sum_{k=1}^N (Total_k - Max(C_1^k, C_2^k, \dots, C_M^k)) \quad (9)$$

Where, N is the number of the obtained intervals under the current discretization scheme and M is the number of categories in the information table. $Total_k$ is the number of instances contained in the k th interval. C_i^k represents the number of instances of the i th category in the

k th interval, $1 \leq i \leq M$, and $Max(C_1^k, C_2^k, \dots, C_M^k)$ is the largest number of instances among all categories in the k th interval.

We use the proposed method to discretize the above data, then compare with EDiRa [35], ChiMerge [34], 1R [1], NCAIC [47], FUDC [50], Cramer's V-Test [43], Chi2 [31], these seven state-of-the-art discretization methods. The results of the number of intervals for each band, indiscernible relationship differences, data inconsistency, and system runtime are shown in Tables 2 and 3.

As shown in Tables 2 and 3, we can see that 1R algorithm obtains the minimum number of breakpoints in the four bands, but the extent of change in the indiscernible relationship is the largest, reaching 12 level, the data inconsistency is also the highest, reaching 38 errors. The extent of change in indiscernible relationship of ChiMerge algorithm is 2 level, and its data errors is two more than our method, but it is 127 more than the number of breakpoints obtained by our method. Although EDiRa algorithm has almost the same number of breakpoints as the proposed method, the extent of change in indiscernible relationship is up to 4 level, and the number of data errors is more than three times that of the method in this paper. NCAIC, Cramer's V-Test and Chi2 have the same degree of change in indiscernible relationship and number of data errors as ChiMerge. Their breakpoints are respectively 215, 87, 100 more than our method. The extent of change in indiscernible relationship of FUDC algorithm is 4 level, its data errors is 2 more than MFD-mvtR, and the number of breakpoints is also 61 more than our method. The seven algorithms have similar performance in terms of running time, however, the proposed method is slightly better. Based on the above analysis, the overall performance of the proposed method is best in the eight algorithms. Figure 4 shows a performance comparison of the eight methods on the number of intervals and data consistency.

The green area in Fig. 4 is the ideal solution range for experimental prediction. In these eight algorithms, only the result produced by our proposed method falls into this ideal region. The experimental results obtained above were analyzed in this paper to find out the reasons as follows. The method of this article is discretized by features grouping at the beginning. After combining the rough set to optimize the result, the number of the indiscernible relationship differences is reduced to 0. So, the minimum interval number and the lowest data error are guaranteed. Although EDiRa algorithm uses entropy to measure the stability of the interval, it is necessary to consider overall similarity between the label rankings in the training set while employing the top-down split strategy of MDLP (Minimum Description Length Principle) [35, 36]. Therefore, when the number of samples increases, the time overhead will increase

Table 2 Comparison of the number of intervals in each band

Method	Band1	Band2	Band3	Band4	Number of intervals
MFD-mvtR	350	350	346	299	1345
EDiRa	349	349	346	299	1343
ChiMerge	442	433	298	299	1472
1R	64	70	42	40	216
NCAIC	447	468	346	299	1560
FUDC	397	439	312	258	1406
V-Test	432	423	288	289	1432
Chi2	410	438	325	272	1445

Table 3 Comparison of performance

Method	Differences	Inconsistencies	Run time
MFD-mvtR	0	5	53 s
EDiRa	4	16	98 s
ChiMerge	2	7	68 s
1R	12	38	56 s
NCAIC	2	7	95 s
FUDC	4	7	85 s
V-Test	2	7	91 s
Chi2	2	7	86 s

significantly. In addition, because it discretizes only one band at a time, the results obtained will destroy the compatibility of the system to some extent. So, we can see that the number of indiscernible relationship differences it obtains is larger than the number of the proposed method which equally use the entropy to measure the intervals. FUDC algorithm also uses entropy to measure the stability of the interval. But unlike EDiRa algorithm, Eq. 1 in the rough set is used by FUDC to define the uncertainty of decision system. Therefore, FUDC has much fewer errors than EDiRa. However, as mentioned in Chapter 1, uncertainty only reflects the number of differences in elements of the equivalence class before and after discretization, and does not represent the number of differences in the equivalence class of decision system, thus, it is not appropriate to use uncertainty to measure the compatibility of decision system. NCAIC algorithm uses class-attribute interdependency as the partitioning criterion of the interval. In addition, the upper approximation of each class and data distribution information are both

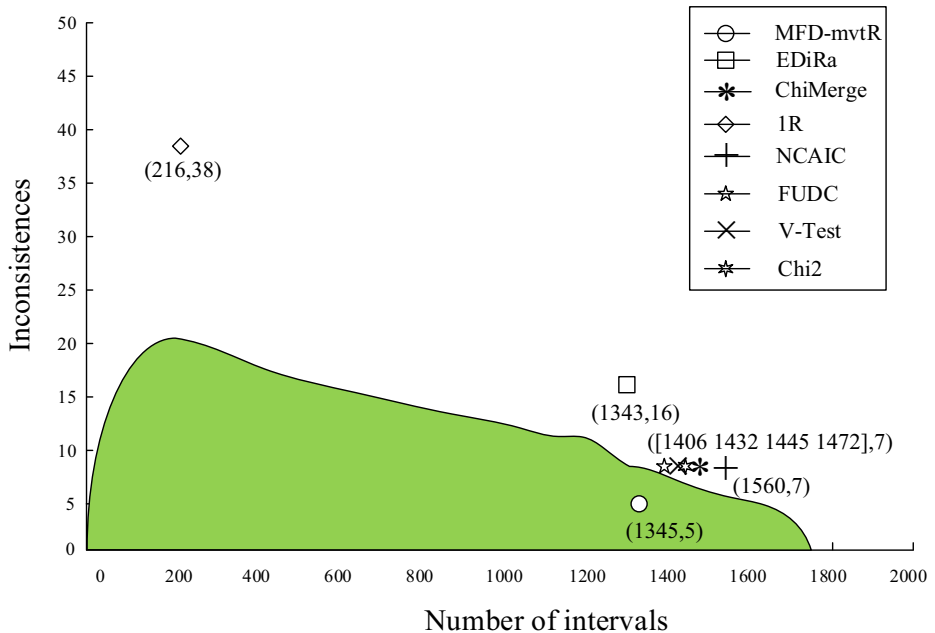


Fig. 4 Comparison of the eight methods on the number of intervals and inconsistencies

considered. However, considering only the upper approximation does not fully describe the entire equivalence class, the discrete discriminant still has a certain probability to skew the class attribute containing the most samples in the interval, resulting in an excessive number of intervals. Therefore, we can see that although NCAIC obtains fewer errors, the number of intervals is the most among the eight algorithms. ChiMerge algorithm uses the method of calculating the category information based on the similarity of intervals to judge and merge adjacent intervals. It use Pearson statistics to determine whether the current breakpoint should be removed, i.e. whether the two intervals adjacent to the breakpoint should be merged. Although it guarantees the mutual exclusion of adjacent intervals, it does not guarantee the stability of categories within an interval. In order to make the interval stability meet the requirements as much as possible, it is necessary to increase the number of intervals as a cost. Therefore, we can see that the number of intervals obtained by ChiMerge is second only to NCAIC. Based on ChiMerge, Cramer's V-Test algorithm weakens the huge influence of n in the discretization scheme through dividing χ^2 by $ln(n)$, where n is the number of intervals. Although the discretization process can be accelerated in some occasions, like ChiMerge, the number of intervals obtained is large because only considering the mutual exclusion of adjacent intervals. Although Chi2 [31] algorithm and Extended Chi2 [16] algorithm proposed later both improve the criteria for determining the importance of breakpoints, the lack of related theoretical evidence still leads to the above discussed problems. The number of intervals for 1R algorithm is given by the user, but the criteria for dividing the interval are too simple and lack flexibility. Although it can quickly obtain the result of discretization, it cannot guarantee both the mutual exclusion of adjacent intervals and the stability of the interior of interval, causing great damage to the compatibility of the system. Therefore, we can see that it has obtained the largest number of indiscernible relationship differences and data errors.

4.4 Evaluation of classification accuracy

The evaluation method at pixel-level is usually adopted for classification accuracy of remote sensing image. This evaluation method is to randomly select the sample data on the classification map then evaluate the classification accuracy by statistically analyzing and comparing with the actual measurement results. The result of the classification accuracy evaluated at pixel-level is usually represented by confusion matrix [23]. The definition of confusion matrix is as follows.

$$CM = (cm_{ij})_{n \times n} = \begin{bmatrix} cm_{11} & cm_{12} & \cdot & \cdot & \cdot & cm_{1n} \\ cm_{21} & cm_{22} & \cdot & \cdot & \cdot & cm_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ cm_{n1} & cm_{n2} & \cdot & \cdot & \cdot & cm_{nm} \end{bmatrix} \quad (10)$$

In the above matrix, n is the total number of categories in the remote sensing image, and cm_{ij} is the number of pixels in the test sample set that should belong to the i th category but are classified into the j th category. Obviously, the greater the value of the diagonal elements in confusion matrix is, the higher the classification accuracy becomes. On the contrary, the smaller value of the diagonal elements in confusion matrix indicates that the number of classification errors is more and the classification accuracy is lower. So, we can get the overall average prediction accuracy through confusion matrix. As shown in Eq. 11.

$$P_{Accuracy} = \frac{\sum_{i=1}^n cm_{ii}}{\sum_{i=1}^n \sum_{j=1}^n cm_{ij}} \tag{11}$$

It is actually the ratio of the number of correctly classified instances to the total number of samples. We can also get the user’s accuracy of the specified category. As shown in Eq. 12.

$$P_u^i = \frac{cm_{ii}}{\sum_{j=1}^n cm_{ij}} \tag{12}$$

Where, P_u^i is the user’s accuracy of the i th category. It is the ratio of the number of correctly classified instances in the i th category to the number of instances contained the i th category. The overall average prediction accuracy and the specified category of user’s accuracy both describe the classification accuracy from different aspects. Their calculations are simple and have a clear statistical significance.

As well as confusion matrix, Kappa coefficient [41] is also widely used in remote sensing image classification accuracy evaluation. Based on confusion matrix, it quantifies the overall effectiveness of the classifier. The expression of Kappa coefficient is shown in Eq. 13.

$$Kappa = \frac{T \sum_{i=1}^n cm_{ii} - \sum_{i=1}^n (cm_{i+} cm_{+i})}{T^2 - \sum_{i=1}^n (cm_{i+} cm_{+i})} \tag{13}$$

Where, T is the total number of pixels used for accuracy evaluation and n is the number of categories. cm_{ii} is the number of pixels on the i th row and i th column in confusion matrix, i.e., the number of correctly classified pixels. cm_{i+} is the total number of pixels on the i th row and cm_{+i} is the total number of pixels on the i th column, respectively. Compared with confusion matrix, Kappa coefficient not only takes account of the correctly classified pixels on the diagonal, but also considers errors of omission and commission that are not on the diagonal. Thus, the two evaluation indicators, confusion matrix and Kappa coefficient, are not equal in general. At present, the application of neural network technology in remote sensing image processing is more and more advanced and comprehensive [12, 15, 33, 45, 48]. It has become

Table 4 The classification accuracy of the eight algorithms

Method	Accuracy	Kappa
MFD-mvtR	85.2083%	0.8225
EDiRa [35]	74.1667%	0.6900
ChiMerge [34]	80.2083%	0.7625
1R [1]	56.4583%	0.4775
NCAIC [47]	80.6250%	0.7675
FUDC [50]	76.2500%	0.7150
V-Test [43]	79.7917%	0.7575
Chi2 [31]	79.3750%	0.7525

an efficient and reliable method for classification of remote sensing images. Table 4 shows the results of the eight algorithms analyzed on the neural network classifier.

It can be seen from the table that the proposed method has the best result in terms of the average prediction precision of the six categories of boats, ports, building, bare land shallow, water body and vegetation, which is about 10% higher than EDiRa algorithm. On the other hand, we can also see that the number of indiscernible relationship differences has a greater impact on the accuracy of the classification. The number of indiscernible relationship differences in ChiMerge algorithm is only 2 fewer than that in the proposed method, but the accuracy is different by 5 percentage points. NCAIC, Cramer's V-Test, Chi2 and ChiMerge are consistent in terms of the number of indiscernible relationship differences, so their accuracies are approximate. Similarly, the accuracies of FUDC and EDiRa are also approximate. 1R algorithm has the largest number of indiscernible relationship differences, so the accuracy obtained is the lowest. We adjust the proposed method parameters to obtain the accuracies for the different number of bands [44] under the different number of indiscernible relationship differences, as shown in Fig. 5.

We can see from Fig. 5, with the increase of the number of bands, the accuracy also rises. Conversely, the increase in the number of indiscernible relationship differences leads to a decrease in accuracy. Figure 6 is a classification effect chart obtained by these eight algorithms in turn.

From (a) to (h) in Fig. 6 correspond to the proposed method, EDiRa algorithm, ChiMerge algorithm, 1R algorithm, NCAIC algorithm, FUDC algorithm, Cramer's V-Test algorithm and Chi2 algorithm, respectively. It can be seen from Fig. 6 that the texture of the objects in Fig. 6a is clearer, and the vessels on the image can be more effectively identified. In particular, the junction between the docking vessel and the port can be well separated. Compared with the classification diagrams of the other seven algorithms, there are fewer bright fringes and the boundaries of each category are clear in the classification diagram of MFD-mvtR algorithm.

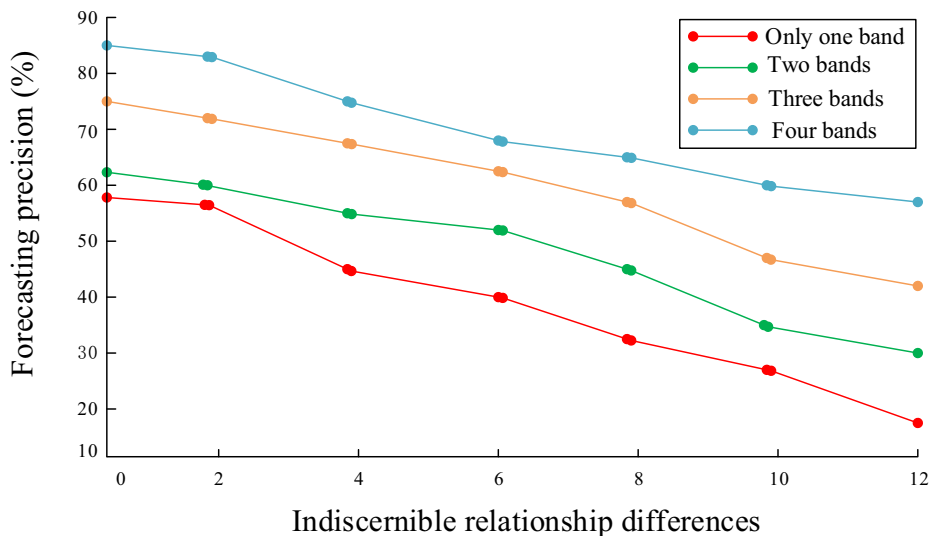


Fig. 5 Effects of feature dimensionality on classification accuracy under different numbers of indiscernible relationship differences

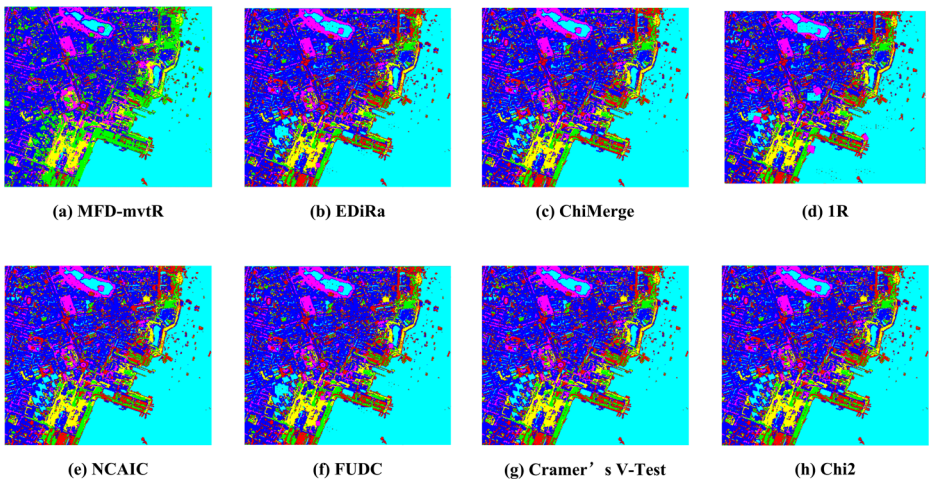


Fig. 6 Classification effect chart of the eight methods

However, the middle areas of (b) to (h) in Fig. 6 respectively have a certain number of bright fringes to different extents, and there are also unrecognizable spots in the water area. Especially in Fig. 6d, the boundary between the docking vessel and the port is blurred, and there are a lot of unrecognizable spots in the water area. From this point of view, the quality of the classification map of our method obtained by the classifier is better than the others. The effect of the proposed method on vessel target recognition is shown in Fig. 7.

Figure 7a is the original remote sensing image, which contains a total of 48 vessels, outlined by red lines. Figure 7b is the classification effect chart of the proposed method, only the ports and the coastline to the sea are highlighted. The detection rate, false alarm rate, and missed alarm rate are measured by the total number of ships [28, 29]. The results of comparative experiments in the proposed method and the other seven algorithms are show in Table 5. As can be seen from Table 5, the discretization results obtained by the proposed method can be applied to the ship target recognition to gain both a higher detection rate and a lower false alarm rate [25]. The comparison result of detection rate is generally consistent with

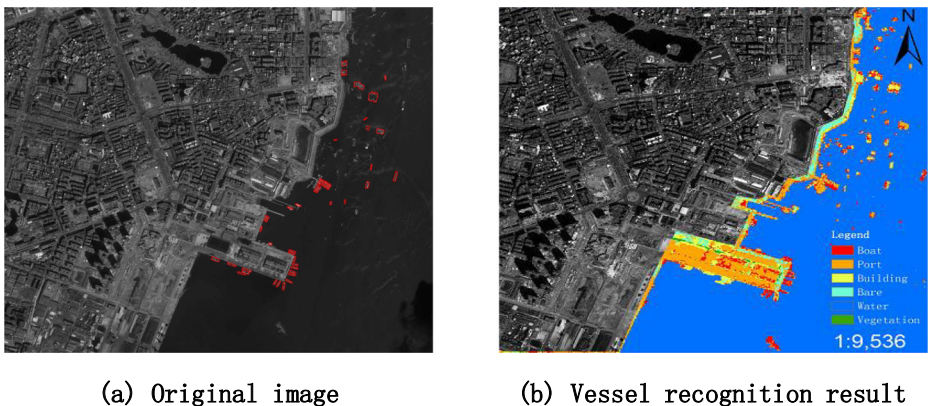


Fig. 7 The result of vessel targets recognition

Table 5 The comparison experiment results of vessels target recognition

Method	Detection rate	False alarm rate	Missed alarm rate
MFD-mvtR	83.3%	7.0%	16.7%
EDiRa [35]	75.0%	12.2%	25.0%
ChiMerge [34]	79.2%	11.6%	20.8%
1R [1]	66.7%	17.9%	33.3%
NCAIC [47]	79.2%	9.5%	20.8%
FUDC [50]	75.0%	10.0%	25.0%
V-Test [43]	77.1%	9.8%	22.9%
Chi2 [31]	77.1%	11.9%	22.9%

that of the previous classification accuracy. It is related to the number of indiscernible relationship differences. NCAIC, Cramer's V-Test, Chi2 and ChiMerge are consistent in terms of the number of indiscernible relationship differences, so their detection rates are approximate. Similarly, the detection rates of FUDC and EDiRa are the same. 1R algorithm has the largest number of indiscernible relationship differences, so, its detection rate is the lowest. Our method benefits from the fact that the level of indiscernible relationship difference can be controlled to zero, thus the highest detection rate is achieved.

5 Conclusions and future work

In this paper, a multivariable optical remote sensing image feature discretization method applied to marine vessel targets recognition is proposed to solve the problem of discretization of marine remote sensing data with multiple features. Firstly, based on the sample set with DN values and labels, an image information decision table which use bands as condition attributes and use land cover classes as decision attribute is established. Secondly, adopt the Top-Down discretization method for each band in the image to calculate the information entropy of all the intervals in the current band, then select the interval with the highest entropy value for splitting. Thirdly, discretize the original decision table by the obtained candidate breakpoints, and introduce the equivalent model of rough set to compare the upper and lower approximation sets of the original decision table with that of the new decision table to get the extent of change in the indiscernible relationship of the image information table. Finally, adjust the algorithm parameters and the segmentation threshold according to the extent of the change in the indiscernible relationship, then rescan each band until termination condition are met to obtain the optimal discretization result. Simulation experiments verify the effectiveness of the proposed method. Compared with other algorithms, it can obtain fewer intervals and higher accuracy. It provides a new idea for preprocessing of optical remote sensing image. It also brings certain guiding significance to the analysis and design of the discretization methods in the marine targets recognition application. Applying our method to other datasets for further testing and improvement is the work to be prepared in the future.

The innovations in this article mainly come from the following aspects: (1) by analyzing the distribution characteristics of DN values of boat and port in each band of the remote sensing images, a basic framework of optical remote sensing image feature discretization for marine vessels target recognition was established, and the original features were grouped to solve the problem that multiple bands interfere with each other in the process of discretization; (2) the compatibility of the system was measured by replacing the γ in the rough set with the number of indiscernible

relationship differences, and information loss after discretization was largely avoided; (3) information entropy was introduced to continuously evaluated for the generated discretized intermediate results, and the feature space was repeatedly scanned to obtain the optimal intervals.

Future research work includes: (1) Apply the method of this article to other data sets (especially high-dimensional data, such as various hyperspectral remote sensing images) for testing and improvement, and expand its scope of utilization to make it more practical; (2) Apply this method to different classifiers for performance comparison and continue to optimize the algorithm model; (3) Test this method in some complex marine environments, so as to continue to perfect its implementation framework for marine targets recognition and detection applications.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant #: 61462022), the National Key Technology Support Program (Grant #: 2015BAH55F04, Grant #:2015BAH55F01), Major Science and Technology Project of Hainan province (Grant #: ZDKJ2016015), Natural Science Foundation of Hainan province (Grant #: 20156235, Grant #: 614232 and Grant#:617062), Higher Education Reform Key Project of Hainan province (Hnjg2017ZD-1), Scientific Research Staring Foundation of Hainan University (Grant #: kyqd1610), the National Natural Science Foundation of China [61762033], the National Natural Science Foundation of Hainan [2018CXTD333, 617048].

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Ali Z, Shahzad W (2016) Comparative Study of Discretization Methods on the Performance of Associative Classifiers. *Frontiers of Information Technology (FIT)*, Islamabad, pp 87–92
2. Catlett J (1991) On changing continuous attributes into ordered discrete attributes. In: *Proceedings of the European Working Session on Learning*. Springer, Berlin, Heidelberg, pp 164–178
3. Chen Z, He C, Zhao C, Xie F (2017) Using SVD-FRFT Filtering to Suppress First-Order Sea Clutter in HFSWR. *IEEE Geosci Remote Sens Lett* 14(7):1076–1080
4. Chen H, Qian C, Zheng H, Wang H (2018) A multilinear unsupervised discriminant projections method for feature extraction. *Multimed Tools Appl* 77(3):3857–3870
5. Cheng J, Xu R, Tang X, Sheng VS, Cai C (2018) An Abnormal Network Flow Feature Sequence Prediction Approach for DDoS Attacks Detection in Big Data Environment. *Computers, Materials & Continua* 55(1):95–119
6. Cheng J, Zhou J, Liu Q, Tang X, Guo Y (2018) A DDoS Detection Method for Socially Aware Networking Based on Forecasting Fusion Feature Sequence. *Comput J*. <https://doi.org/10.1093/comjnl/bxy025>
7. Di W, Zhang Y, Wang H, Huang M, Feng W, Chen R (2017) Study on the assessment method of typhoon regional disaster based on the change of chlorophyll-a concentration in seawater. *OCEANS 2017, Aberdeen*, pp 1–7
8. Fayyad UM, Irani KB (1992) On the Handling of Continuous-Valued Attributes in Decision Tree Generation. *Mach Learn* 8(1):87–102
9. Grzymalabuse JW, Mroczek T (2016) A Comparison of Four Approaches to Discretization Based on Entropy. *Entropy* 18(3):69
10. Halvor S, Bakkeløkken HK, Hao W, Ottar O (2017) Measuring Container Port Complementarity and Substitutability with Automatic Identification System (AIS) Data – Studying the Inter-port Relationships in the Oslo Fjord Multi-port Gateway Region. *TransNav, International Journal on Marine Navigation and Safety of Sea Transportation* 11(2):79–84
11. He X, Min F, Zhu W (2014) Comparison of Discretization Approaches for Granular Association Rule Mining. *Can J Electr Comput Eng* 37(3):157–167
12. Heermann PD, Khazenie N (1992) Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Trans Geosci Remote Sens* 30(1):81–88
13. Jin R, Breitbart Y, Muoh C (2009) Data discretization unification. *Knowl Inf Syst* 19(1):1–29

14. Kaur J, Kaur K (2017) A Fuzzy Approach for an IoT-based Automated Employee Performance Appraisal. *Computers, Materials & Continua* 53(1):23–36
15. Kumar DA, Meher SK, Kumari KP (2017) Knowledge-Based Progressive Granular Neural Networks for Remote Sensing Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10(12):5201–5212
16. Lavagnananda K, Chattanachot S (2017) Study of discretization methods in classification. *Knowledge and Smart Technology (KST), Chonburi*, pp 50–55
17. Lee C, Tsai C, Yang Y, Yang W (2007) A Top-Down and Greedy Method for Discretization of Continuous Attributes. *Fuzzy Systems and Knowledge Discovery, Haikou, China*, 1:472–476
18. Li W, Fu K, Sun H, Sun X, Guo Z, Yan M, Zheng X (2017) Integrated Localization and Recognition for Inshore Ships in Large Scene Remote Sensing Images. *IEEE Geosci Remote Sens Lett* 14(6):936–940
19. Liu H, Hussain F, Tan CL, Dash M (2002) Discretization: An Enabling Technique. *Data Min Knowl Disc* 6(4):393–423
20. Liu H, Liu D-Y, Shi X-H, Gao Y (2008) An attribute discretization algorithm based on Rough Set and information entropy. *Machine Learning and Cybernetics, Kunming*, pp 206–211
21. Ma Y, Luo X, Li X, Bao Z, Yi Z (2018) Selection of Rich Model Steganalysis Features Based on Decision Rough Set α -Positive Region Reduction. *IEEE Transactions on Circuits and Systems for Video Technology*. <https://doi.org/10.1109/TCSVT.2018.2799243>
22. Morente-Molinera JA, Mezei J, Carlsson C, Herrera-Viedma E (2017) Improving Supervised Learning Classification Methods Using Multigranular Linguistic Modeling and Fuzzy Entropy. *IEEE Trans Fuzzy Syst* 25(5):1078–1089
23. Ohsaki M, Wang P, Matsuda K, Katagiri S, Watanabe H, Ralescu A (2017) Confusion-Matrix-Based Kernel Logistic Regression for Imbalanced Data Classification. *IEEE Trans Knowl Data Eng* 29(9):1806–1819
24. Pai GAV (2017) Fuzzy Decision Theory Based Metaheuristic Portfolio Optimization and Active Rebalancing Using Interval Type-2 Fuzzy Sets. *IEEE Trans Fuzzy Syst* 25(2):377–391
25. Patel V, Madhukar H, Ravichandran S (2018) Variability index constant false alarm rate for marine target detection. *Signal Processing and Communication Engineering Systems (SPACES), Vijayawada*, pp 171–175
26. Patra S, Modi P, Bruzzone L (2015) Hyperspectral Band Selection Based on Rough Set. *IEEE Trans Geosci Remote Sens* 53(10):5495–5503
27. Pawlak Z (1992) *Rough Set: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Norwell
28. Qi S, Ma J, Lin J, Li Y, Tian J (2015) Unsupervised Ship Detection Based on Saliency and S-HOG Descriptor From Optical Satellite Images. *IEEE Geosci Remote Sens Lett* 12(7):1451–1455
29. Qi S, Ma J, Tao C, Yang C, Tian J (2013) A Robust Directional Saliency-Based Method for Infrared Small-Target Detection Under Various Complex Backgrounds. *IEEE Geosci Remote Sens Lett* 10(3):495–499
30. Qingyao W, Ye Y, Liu Y, Ng MK (2012) SNP Selection and classification of Genome-Wide SNP Data Using Stratified Sampling Random Forests. *IEEE Transactions on NanoBioscience* 11(3):216–227
31. Qu W, Yan D, Sang Y, Liang H, Kitsuregawa M, Li K (2008) A novel Chi₂ algorithm for discretization of continuous attributes. *10th Asia Pacific Web Conference, Shenyang*, pp 560–571
32. Ramirezgallego S, Garcia S, Mourintalin H, Martinezregio D, Boloncanedo V, Alonsobetanzos A, Benitez JM, Herrera F (2016) Data discretization: taxonomy and big data challenge. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery* 6(1):5–21
33. Ren L, Sun Y, Wang H, Zhang L (2018) Prediction of Bearing Remaining Useful Life with Deep Convolution Neural Network. *IEEE Access* 6:13041–13049
34. Rosati S, Balestra G, Giannini V, Mazzetti S, Russo F, Regge D (2015) ChiMerge discretization method: Impact on a computer aided diagnosis system for prostate cancer in MRI. *IEEE International symposium on Medical Measurements and Applications (MeMeA) Proceedings, Turin*, pp 297–302
35. Sá CR, Soares C, Knobbe A (2016) Entropy-based discretization methods for ranking data. *Inf Sci* 329(1):921–936
36. Garcia S, Luengo J, Sáez JA, López V, Herrera F (2013) A Survey of Discretization Techniques: Taxonomy and Empirical Analysis in Supervised Learning. *IEEE Trans Knowl Data Eng* 25(4):734–750
37. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
38. Stender DHS, Berg H, Hjelmeryk KT, Sæstad TS (2017) The classification performance of signal-to-noise ratio and kinematic features in varying environments. *OCEANS, Aberdeen*, pp 1–5
39. Sun Y, Yang Y, Li Y, Zhang Y (2015) Full diversity reception based on Dempster-Shafer theory for network coding with multiple-antennas relay. *China Communications* 12(10):76–90
40. Tao X, Duan Y, Ge N (2017) K-NN based bypass entropy and mutual information estimation for incremental remote-sensing image compressibility evaluation. *China Communications* 14(8):54–62
41. Vieira SM, Kaymak U, Sousa JMC (2010) Cohen's kappa coefficient as a performance measure for feature selection. In: *Fuzzy Systems (FUZZ)*. IEEE, Barcelona, pp 1–8
42. Wang C, Xu Z, Wang S, Zhang H (2018) Semi-supervised classification framework of hyperspectral images based on the fusion evidence entropy. *Multimed Tools Appl* 77(9):10615–10633

43. Wu B, Zhang L, Zhao Y (2014) Feature Selection via Cramer's V-Test Discretization for Remote-Sensing Image Classification. *IEEE Trans Geosci Remote Sens* 52(5):2593–2606
44. xie L, Li G, Xiao M, Peng L (2016) Novel classification method for remote sensing images based on information entropy discretization algorithm and vector space model. *Comput Geosci* 89(C):252–259
45. Xu X, Li W, Ran Q, Du Q, Gao L, Zhang B (2018) Multisource Remote Sensing Data Classification Based on Convolutional Neural Network. *IEEE Trans Geosci Remote Sens* 56(2):937–949
46. Xu Y, Ma P, Ng E, Lin H (2015) Fusion of WorldView-2 Stereo and Multitemporal TerraSAR-X Images for Building Height Extraction in Urban Areas. *IEEE Geosci Remote Sens Lett* 12(8):1795–1799
47. Yan D, Liu D, Sang Y (2014) A new approach for discretizing continuous attributes in learning systems. *Neurocomputing* 133(10):507–511
48. Yıldız SA, Öztürk AU (2016) A Study on the Estimation of Prefabricated Glass Fiber Reinforced Concrete Panel Strength Values with an Artificial Neural Network Model. *Computers, Materials & Continua* 52(1):41–52
49. Zeng D, Dai Y, Li F, Simon Sherratt R, Wang J (2018) Adversarial Learning for Distant Supervised Relation Extraction. *Computers, Materials & Continua* 55(1):121–136
50. Zhang G, Wu Z, Yi L (2008) A Remote Sensing Feature Discretization Method Accommodating Uncertainty in Classification Systems. *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, pp 195–202
51. Zheng R, Xi G (2009) The application of discretization based on rough set and information entropy in TCM. *Nature & Biologically Inspired Computing*, Coimbatore, pp 1695–1701

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Mengxing Huang is currently a Professor and PhD supervisor of computer science and technology, and the Dean of College of Information Science and Technology. He is also the Vice-President of the state key laboratory marine resource utilization in South China Sea. He received his PhD from Northwestern Polytechnical University in 2007. He then joined staff at the Research Institute of Information Technology in Tsinghua University as a postdoctoral researcher. His research interests include big data and intelligent information processing, cloud computing and internet of things (IoT), etc. He is a member of IEEE, and a senior member of Chinese Computer Federation (CCF), and a member of the Information System Committee in CCF.



Qiong Chen received his B.Eng. degree at Beijing University of Posts and Telecommunications, P.R.China, 2007 and M.Eng. degree at Politecnico di Torino, Italy, 2012. Now he is a Ph.D. student at College of Information Science and Technology, Hainan University, P.R.China. His research interests include remote sensing image processing, artificial intelligence and target recognition, evolutionary computing, big data analytics and multivariate data fusion.



Hao Wang is an associate professor and the head of Big Data Lab at the Department of ICT and Natural Sciences in Norwegian University of Science & Technology, Norway. He has worked as a researcher in IBM Canada, McMaster, and St. Francis Xavier University before he moved to Norway. He received a Ph.D. degree in 2006 and a B.Eng. degree in 2000, both in computer science. His research interests include big data analytics and industrial internet of things, high performance computing, safety-critical systems, and communication security. He has published 70+ papers in international journals and conferences such as IEEE TVT, GlobalCom. He served as a TPC co-chair for IEEE DataCom 2015, IEEE CIT 2017, ES 2017 and reviewers for journals such as IEEE TKDE, TBD, TETC, T-IFS, ACM TOMM. He is a member of IEEE IES Technical Committee on Industrial Informatics.