Proceedings of the 8th International Symposium on
Spatial Accuracy Assessment in Natural Resources and Environmental Sciences
Shanghai, P. R. China, June 25-27, 2008, pp. 195-202

A Remote Sensing Feature Discretization Method Accommodating Uncertainty in Classification Systems

Guifeng Zhang, Zhaocong Wu and Lina Yi

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

Abstract. Most of the classification methods in remote sensing can only process the discrete feature data, such as rough set. Thus the discretization of feature plays a very important role in the remote sensing imagery classification system. In general, the remote sensing features are discretized currently by means of the methods from fields other than remote sensing. Because there is a lack of consideration of uncertainty of the classification system in these methods, it is not predicted whether the discretization influences the classification accuracy or not. This paper introduces a discretization method considering the uncertainty of the classification system. It comprises three components: the building of the initial candidate cut points set, the selection of cut points based on information entropy and the deletion of redundant cut points. All the three parts are executed iteratively. The first two are iterative processes from top to bottom, while the last is an iterative process from bottom to top. The stopping criterion of iterative process is a threshold which represents the max possible change of the uncertainty of the classification. Therefore, the change of the uncertainty of the classification system resulted from discretization is limited to a certain range, and its influence on classification can be predicted and controlled. The experiment shows that the proposed method produces comparative results with those of Ent-MDLC and can lessen the influence on the classification accuracy from the discretization.

Keywords: uncertainty, discretization, segment range, rough entropy, accuracy

1. Introduction

With the advancement of remote sensing intelligent processing technology, the features involved in remote sensing classification are various, including spectral features, texture features, and geometrical features etc. These features can be divided into two groups: continuous features and discrete features. However, some of the effective classification methods can only process the discrete features, such as rough set. The discretization of features plays an important role in the pre-process of classification. The discretization of features benefits the classification system in two sides. Firstly, an effective discretization can highly improve the cluster ability of training samples and enhance the robustness of the classification system. Secondly, the discretization can reduce the attributes involved and simplify the structure of the classification system. Therefore, the discretization of features plays an important role in the pre-process of classification.

In different study field, there are numbers of data discretization methods, which can be classified according to five axes: supervised versus unsupervised, dynamic versus static, global versus local, splitting versus merging, direct versus incremental [1], [2]. The classification information of samples is taken into account in the supervised discretization methods, while it is not used in the unsupervised discretization methods, such as equal-width and equal-frequency. In the dynamic discretization methods, the discretization is synchronous with the building of the classifier. On the contrast, in the static methods, it is done before the building of the classifier. Local approaches discretize each attribute independently while global approaches discretize all attributes simultaneously. Splitting and merging discretization methods are different in the searching scheme. In splitting discretization the top-down scheme is used while in merging discretization the bottom-up scheme is used. The direct and incremental methods are different in the setting of number of cut-points. In the direct methods, the number of cut-points is set before discretization, but in the incremental

ISBN: 1-84626-171-6, **ISBN13:** 978-1-84626-171-8

methods the number of cut-points changes with the improvement of the discretization results until the discretization result meets certain requirements.

In general, the discretization methods currently used in classification are from fields other than Remote Sensing. Although these methods meet the requirements of discretization of remote sensing feature data to some extent, they are lack of considering the character of the remote sensing feature data. Without considering the uncertainty of the classification system of remote sensing data, the classification precision of the system is possible to decrease after discretization. It is not predicted how the change of uncertainty will influence the classification precision in the classification system For the purpose of reducing the influence of the data discretization on the classification accuracy, the uncertainty change of classification system should be taken into count in discretization.

A discretization method considering the uncertainty of the system of classification is proposed in this paper, which can reduce the influence produced by the discretization. The cut points are generated in a top-down iterative computation, its stopping condition is defined by the uncertainty change of the classification system. The investigation of this method in the paper mainly consists of four parts: representation of the uncertainty of classification system (section 2); the discretization method for the remote sensing feature data considering the uncertainty of the classification system (section 3); comparison of the classification results of the original remote sensing data and discretized data by the decision tree (section 4); conclusion (section 5).

2. Representation of the uncertainty of classification system

The quality of discretization involves a tradeoff between simplicity and the uncertainty change. If the uncertainty change can be properly controlled, the discretization may achieve predicted results. Therefore, this paper investigates a discretization method considering the uncertainty of the classification system. For the purpose of controlling the uncertainty change of the discretized classification system, once the cut point set changes, this method computes the uncertainty and judges whether the uncertainty change is less than a limited value or not. Therefore, this paper introduces the uncertainty measurement of the decision table based on rough set [3]. This measure can reflect the classification accuracy of the classification system. But it is not sensitive to the change of information in the classification while a cut point is inserted into or removed from the cut point set. Then, this paper uses another measure of uncertainty—the partition entropy, which measures the change of uncertainty when the sample set is divided into several subsets.

2.1. Uncertainty measurement of decision table based on rough set

In this paper, the classification system is represented as a decision table which the condition attributes are the feature of samples and the decision attribute is the category of the samples.

When there is not conflict in the decision table, the decision table is certain. In contrast, if some of the classification isn't consistent, the information and knowledge acquired from it can not be totally certain, it is uncertain. The uncertainty of the decision table can be defined as follow.

In a information expression system represented by $S = < U, R, V, f >, R = C \cup D$, where C is the condition attribute set and D is the decision attribute set, U denote a finite and non-empty set of objects called the universe and V, f denote the value range of C and D that are equivalence relations on U. Let E_i be an equivalence class of C which is represented by $E_i \in U \mid IND(C)$ where $IND(C)$ is the indiscernible relation, $i = 1,2,......m$ and m is the class number. Let X_i be an equivalence class of D, which is represented by $X_i \in U \mid IND(D)$, where $i = 1,2,......n$ and n is the class number. The certainty degree of E_i can be calculated by formula 1.

$$\mu_{\max}(E_i) = \max(\{| E_i \cap X_j | / | E_i | : X_j \in U \mid IND(D)\}) \tag{1}$$

So the uncertainty of the decision table is defined in formula 2.

$$\mu_{uncer}(S) = 1 - \sum_{i=1}^{m} \frac{| E_i |}{| U |} \bullet \mu_{\max}(E_i) \tag{2}$$

It is can be seen that this measure of the uncertainty is equal to the ratio of the number of instances that can't be classified correctly to the number of all instances. Therefore, it can express the conflict degree of the decision table.

Besides, it is obvious that $0 \le \mu_{uncer}(S) \le 1$. If there isn't conflict in the decision table, the uncertainty value is 0. If all the classification is contradictory, the uncertainty value is 1.

2.2. The entropy of the set which is divided into several subsets

If the classes are regarded as the possible event in the classification system, we can achieve another expression of the uncertainty. Before describing this expression, the Shannon entropy is introduced firstly. It is defined in formula 3.

$$H = H(p_1, p_2 ... p_m) = -\sum_{i=1}^{m} p_i \log_2 p_i \tag{3}$$

Where, p_i is the possibility of the ith event.

If we define the possibility of the class as $P_i = |X_i| / |S|$, the entropy of the classification system can be defined in formula 4.

$$H(S) = H(X_1, X_2 ... X_m) = -\sum_{i=1}^{m} \frac{|X_i|}{|S|} \log_2 \frac{|X_i|}{|S|} \tag{4}$$

Where, m is the class number of samples.

X_i - is the set of which elements belong to the i th class.

S – is the sample set,

$|X_i|$ - is the elements number of the X_i ,.

In the discretization, when the set is divided into several subsets, the entropy of the divided set should be equal to the weight sum of the entropy of several subsets. Therefore, the entropy of the sample set that are divided into l subsets is defined in the formula 5.

$$H(T,S) = \frac{|S_1|}{|S|} H(S_1) + \frac{|S_2|}{|S|} H(S_2) + ... + \frac{|S_l|}{|S|} H(S_l) \tag{5}$$

Where, S_i is the ith subset. This measure is named as partition entropy in this paper.

If the cut point set is changed, we compute the difference of the partition entropy of the samples discretized by cut point before and after change. This difference is named as information gain, and it can express the change value of information from the change of the cut point set.

3. Discretization of the remote sensing feature data considering the uncertainty of the classification system

According to the two expressions of the uncertainty, the discretization method proposed by this paper follows three principles.

- Once the cut points set changes, the method controls the iterative computation in accordance with the uncertainty change. In this case, the uncertainty measure is computed based on rough set.
- During judging which a point is added into the cut point set or which a point is removed from the cut point set, the information gain is defined as the criterion. The point with max information gain is accepted as a new cut point and the point with min information gain is accepted as a candidate point which will be removed.

Therefore, our discretization method has the following steps.

- Building the initial cut points set. For each attribute, the attribute distributed ranges of each class are decided firstly. Then we can achieve a series of end points of the ranges, and these points are regarded as the initial cut point set of this attribute. After all the attributes are processed, the uncertainty change is computed and judged. If it exceeds a limited value, the computation continues. Otherwise, the computation stops.
- Selecting cut point based on entropy. If a point is added into the cut point set and the divided set entropy achieves max value, the method accepts it as a cut point. After new points being inserted, if the uncertainty of the classification system is less than a limited value, this step will be terminated.
- Reducing the redundant cut points. If the divided set achieves max entropy after one of the cut points is removed from the cut point set and the uncertainty is less than a limited value, this point is decided

as a redundant cut point and removed from the cut point set. After all the redundant points are removed from the cut point set, this step stops.

In the following, this paper describes the three steps in detailed.

3.1. The building of initial cut points set

The building of initial cut points set is an iterative process. It consists of the following steps.

1. Setting multiple $n_l = n_{l-1} + dn$ in the l th iterative step, where dn is the adjusting step length of the multiple and l is the iterative sequential number. In the first iterative step, n_0 and dn are initialized. The two parameters control the min distance between two attribute ranges, and they should be initialize according to the data set.

2. Creating the initial cut point set. Here the cut point set of each attribute is computed one by one.

In the process of creating the cut point set of attribute a (a denotes the attribute number), we compute the attribute ranges of each class. During computing the attribute a ranges of the m class (m denotes the class number), there exist the following steps.

- Creating the value set of attribute i of class m. Here, this value set is being denoted $V = \{v_0, v_1, v_2, \cdots v_h\}$ (h denotes the number of attribute i of class m.). And the elements of V are sorted in an ascending order.
- Computing the difference between the neighbor elements in V and building difference value set $D = \{d_1, d_2 \cdots d_{h-1}\}$. After the standard deviation s_{mi} and the mean value u_{mi} of D are calculated, we can calculate the threshold which denotes the min distance of two attribute value range. And it is defined in formula 6.

$$d\varphi_l = u_{mi} + s_{mi} \times n_l \tag{6}$$

For the two continuous steps, $n_l = n_{l-1} + dn$, where dn is the adjusting step length of the multiple. These are set in the 1 step.

- The first element of V is taken as the beginning value (Vb) and the current value (Vc), and the first element of V is taken as the current difference value (Dc).
- The absolute of Dc is compared with $d\varphi_l$. If the absolute value of Dc is bigger than $d\varphi_l$, Vc is taken as the ending value (Ve), and a new attribute range ([Vb, Vc]) can be achieved. At the same time, the next element of the attribute set is taken as Vc and Vb, and the next element of the difference value set is taken as Dc. Otherwise, the new distribution range is not established and the beginning value is not renewed. We repeat the (4) step until all elements of D are compared.

3. After initial cut point set of all attributes are found out, the classification can be discretized. Then the uncertainty change can be calculated by formula 7.

$$ch_l = \left|\mu_{uncer}(S_s) - \mu_{uncer}(S_l)\right| \tag{7}$$

Where, S_0- is the original classification system,

S_l- is the classification discretized by the cut point set achieved in the l iterative step.

If ch_l is smaller than Thr_{space}, the iterative computation will be stopped. Otherwise the steps 1,2,3 are repeated.

3.2. Selection of cut points based on rough entropy

In the building of the initial cut points set, Thr_{space} is set bigger to improve the efficiency of the algorithm. Therefore, the cut points set got from the 3.1 can not meet the predicted classification accuracy requirement. Thence, new cut points should be selected and inserted into the cut point set.

By adding new cut points into the cut points set continuously, the uncertainty of the classification system will decrease. When the uncertainty change is less than Thr_{table}, the iteration stops, and the classification accuracy meet the requirement. Thr_{table} is the max changed value of the uncertainty of the discretized decision table. Here, the cut points are selected based on information entropy.

Each step of the selection of cut points based on information entropy mainly consists of three steps: the building of inconsistent samples set of each iterative step, the selection of the optimal cut points for each inconsistent samples set, and the computation control.

1. The building of $ICSS_{lm}$.

From the definition of uncertainty of classification system, the uncertainty of classification system can be reduced by eliminating inconsistency. After discretizated, the samples of which all the condition attribute values are same are grouped as a set. If there are inconsistent samples in the set, it is defined as an inconsistent samples set ($ICSS_{lm}$, where l is the iterative number, and m is the sequential number of the set).

2. Selecting new cut points of each $ICSS_{lm}$. In this algorithm, we select the new cut point from $ICSS_{lm}$ orderly. In each $ICSS_{lm}$, we also select the new cut point of each attribute orderly. The selection of the optimal cut points of the jth condition attribute in $ICSS_{lm}$ is mainly consisted of the following steps:

- The first step is to create the value set $V_j = \{v_0, v_1, \cdots, v_{n-1}\}$ of the jth condition attribute value in $ICSS_{lm}$, where n is the number of different jth condition values in $ICSS_{lm}$.
- Secondly, every $v_i \in V_j$ ($i \neq 0, i \neq n-1$) is processed as follows. It is selected as a candidate point T_l by which $ICSS_{lm}$ is divided into two subsets. And the information entropy of the divided $ICSS_{lm}$ can be calculated, which is represented by $H(T_l, ICSS_{lm})$. Then the difference value between every the divided set and the original (information gain) is calculated by formula 8.

$$dH_i = H(ICSS_{lm}) - H(T_i, ICSS_{lm}) \tag{8}$$

- Thirdly, $v_i \in V_j$ with the maximum dH_i is chosen as the new cut point and added into the cut points set of this condition attribute.
- Lastly, the original classification system is discretized by the new cut points set. If the uncertainty change is less than Thr_{table}, the iteration stops. Otherwise, the next inconsistent sample set is processed.

3. The control of the iterative computation. If the discretization result can not reach the requirement for classification accuracy after all the inconsistent sample sets are processed, the step 1 and 2 are repeated. Otherwise, the selection of cut points stops.

3.3. Deletion of the redundant cut points

After the previous two processes, redundant cut points may exist in the acquired cut points set. To get high quality cut points set, the redundant cut points should be deleted. In the deletion of the redundant cut points, the testing method is used widely. In this method, we delete one cut point in the cut point set, then calculate the uncertainty of the discretized decision table by current cut point set. If the uncertainty does not change, this cut point is a redundant point and should be deleted, otherwise it should be reserved. If all the cut points are tested by this method, it will cost a long time. Therefore, the heuristic method is used for the deletion of redundant points in this paper.

In the heuristic method, the cut points set of each attribute is processed successively. For each attribute, the method finds out the cut point with the highest possibility firstly. If it can be deleted, the method deletes it and search the next. Otherwise, if it can not be deleted, the process of this attribute should be ended. The deleted possibility of each cut point is very important in the heuristic redundant cut point deletion method. Here information gain is used to measure this possibility. As the definition of information gain mentioned in section 2, the less information gain after deleting one cut point, the more possible the cut point is redundant. The deleted possibility (information gain) of cut point is calculated by formula 9.

$$P_i = H(T_i, S) - H(T\ , S) \tag{9}$$

Where, $H(T_i, S)$is the information entropy of the system after deleted the ith cut point.

$H(T, S)$is the information entropy of the system before deleting the cut point. The smaller P_i is, the larger the deleting possibility of the ith point is.

4. Experiment results

For the purpose of testing the validity of the method proposed by this paper, two experiments are presented. At first this method and Ent-MDLP are used to discretize the data sets, and the experiment results are analyzed. We get various discretized results by setting different limited uncertainty change in the experiment, and analyze the influence of the change of the uncertainty to the classification accuracy.

Because the Ent-MDLP algorithm has been accepted as one of the best discretization methods, so it is chosen as the comparative target of the method proposed by this paper, and the two methods are used to discretize the data sets got from UCI repository. The discretized results are list in the Table 1 that shows the information of data sets and the number of intervals of data sets produced. It is can be seen from the table that the numbers of intervals of FUDC and Ent-MDLP are approximate. The uncertainty of the data sets doesn't change after discretization by FUDC. But the Ent-MDLP method doesn't consider the uncertainty, the uncertainty of data set after discretization may change. In addition, FUDC can achieve various results by changing the limited change of uncertainty. Therefore, it can be concluded that FUDC performs better than Ent-MDLP.

Table 1. Data sets description and the comparison of number of intervals

Data Sets	Instance number	Attribute Number	Ent-MDLP	FDCU	
			Intervals Number	Uncertainty Change	Intervals Number
Glass	214	9	21	0	24
Ionosphere	351	33	114	0	107
Iris	150	4	10	0	13
Wdbc	569	13	82	0	84
Vehicle	845	30	63	0	58

In the second experiment, an SPOT5 multi-spectral image locating at Jianghan plain is used. It is shown in Figure 1. There are 4 bands in this image. And seven classes of objects are chosen for test, including: water, resident, vegetation and four types of farmlands. The point number of the training set and checking set of each class is listed in Table 2.

The C5.0 decision is chosen as the classifier in the experiment. In the experiment, different thresholds are set in the beginning. The experiment is executed as the following steps: 1.Setting the threshold; 2. Discretizing the sample set and the testing set; 3. Classifying the data by C5.0 decision tree; 4. Evaluating the classification results. Table 2 shows the results of the experiment and there are the numbers of cut points, rules and the classification accuracy in different threshold. From the table, it can be seen that as the change value of the uncertainty increases the number of cut points and rules decrease, and the structure of the decision tree becomes simple. In addition, when the limited changed value of uncertainty is in a certain range, the classification accuracy is in a certain range. However, if the limited changed value of the uncertainty exceeds a certain value, the classification accuracy decreases rapidly. For example, when the limited changed range of the uncertainty exceeds 0.008, the classification accuracy rapidly decreases and is lower than 92.45%. Therefore, in the discretization of remote sensing imagery features, the suitable threshold should be chosen to get less cut points and rules in the condition that the classification accuracy is guaranteed.

Table 2 Points number of training set and checking set

Category	Vegetation	Land1	Land2	Land3	Land4	Water	Resident
Training Set	430	760	658	563	565	600	661
Testing Set	362	448	269	382	472	300	281

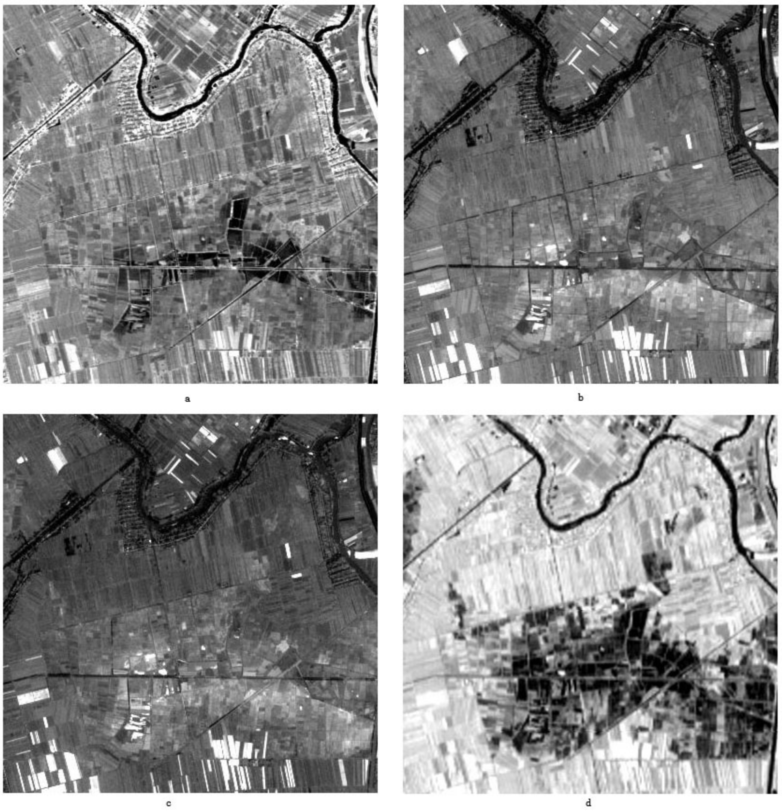

Fig 1: SPOT 5 images: red band (c), green band (b), blue (a), near infrared blue (d)

Table 3. Classification accuracy

Uncertainty Change	Intervals Number	Rules Number	Kappa oefficient	Overall Accuracy（%）
0	83	25	0.9746	97.83
0.001	72	24	0.9423	95.13
0.002	68	23	0.9373	94.75
0.003	63	21	0.9353	94.53
0.004	41	19	0.9448	95.34
0.005	33	18	0.9513	95.89
0.006	32	18	0.9393	94.87
0.007	32	18	0.9498	95.76
0.008	32	18	0.9498	95.76
0.009	29	17	0.9109	92.45
0.01	28	17	0.8857	89.56
0.011	26	16	0.8314	85.65
0.012	23	16	0.8265	83.43
0.013	22	15	0.7754	79.34

5. Conclusions

This paper presents an innovative approach for the discretization of remote sensing feature considering the uncertainty of the classification system, and it is a supervised, static and global method. The top-down iterative algorithm is used in the method to insert new cut points into the cut point set step by step, and the change value of the uncertainty of the classification system is the stopping criterion of the algorithm. The algorithm gets a trade off between simplicity and predictive accuracy. For the purpose of evaluating the performance of the method, it is compared with Ent-MDLP, which is known as one of the best discretization methods [3]. The experimental results show that the bigger the uncertainty change value is, the lower the classifying accuracy is. And a balance between the uncertainty change value and the classification accuracy in the discretization can be achieved by control.

6. Acknowledgements

This paper is supported by the National Natural Science Foundation (No.40201039 and No. 40771157) and the National High Technology Research and Development Program of China (No.2007AA12Z413).

7. References

[1] Xiaoyan Liu and Huaiqing Wang, A Discretization Algorithm Based on a Heterogeneity Criterion, IEEE Trans. Knowledge and Data Eng., 2005, **17**(9): 1166-1173.

[2] Xiaohang Zhang and Jun Wu, “A Discretization Algorithm Based on Gini Criterion”, Sixth International Conference on Machine Learning and Cybernetics, 2007: 19-22.

[3] D.V. Popel and E.I. Popel, “Controlling uncertainty in discretization of continous data”, Multiple-Valued Logic. Proc. 34th Int’l Symp., 2004: 288-293.

[4] Duntsch I, Gediga G, “Uncertainty Measures of Rough Set Prediction”, J of Artificial Intelligence, 1998, **106**(2): 77-107.