Chapter 2
Attribute Selection Based on Reduction of Numerical Attributes During Discretization

Jerzy W. Grzymała-Busse and Teresa Mroczek

Abstract Some numerical attributes may be reduced during discretization. It happens when a discretized attribute has only one interval, i.e., the entire domain of a numerical attribute is mapped into a single interval. The problem is how such reduction of data sets affects the error rate measured by the C4.5 decision tree generation system using ten-fold cross-validation. Our experiments on 15 numerical data sets show that for a Dominant Attribute discretization method the error rate is significantly larger (5% significance level, two-tailed test) for the reduced data sets. However, decision trees generated from the reduced data sets are significantly simpler than the decision trees generated from the original data sets.

Keywords Dominant attribute discretization · Multiple scanning discretization · C4.5 Decision tree generation · Conditional entropy

2.1 Introduction

Discretization based on conditional entropy of the concept given the attribute (feature) is considered to be one of the most successful discretization techniques [1–9, 11, 12, 15–17, 19–22]. During discretization of data sets with numerical attributes some attributes may be reduced, since the entire domain of the numerical attribute is

J.W. Grzymała-Busse (✉)
Department of Electrical Engineering and Computer Science, University of Kansas, 66045 Lawrence, KS, USA
e-mail: jerzy@ku.edu

J.W. Grzymała-Busse · T. Mroczek
Department of Expert Systems and Artificial Intelligence, University of Information Technology and Management, 35-225 Rzeszów, Poland
e-mail: tmroczek@wsiz.rzeszow.pl

U. Stańczyk et al. (eds.), *Advances in Feature Selection for Data and Pattern Recognition*, Intelligent Systems Reference Library 138,
https://doi.org/10.1007/978-3-319-67588-6_2

mapped into a single interval. A new numerical data set may be created by removing attributes from the original numerical data set indicated by single-intervals. Such data sets are called reduced. Our main objective is to compare quality of numerical data sets, original and reduced, using the C4.5 decision tree generation system. To the best of our knowledge, no similar research was ever conducted.

We conducted a series of experiments on 15 data sets with numerical attributes. All data sets were discretized by the Dominant Attribute discretization method [12, 14]. In Dominant Attribute method, first the best attribute is selected (it is called the Dominant Attribute), a then for this attribute the best cutpoint is chosen. In both cases, the criterion of quality is the minimum of corresponding conditional entropy. New, reduced data sets were created. For pairs of numerical data sets: original and reduced, the ten-fold cross-validation was conducted using C4.5 decision tree generation system. Our results show that the error rate is significantly larger (5% significance level, two-tailed test) for the reduced data sets. However, decision trees generated from the reduced data sets are significantly simpler than the decision trees generated from the original data sets. Complexity of decision trees is measured by the depth and size.

2.2 Dominant Attribute Discretization

An example of a data set with numerical attributes is presented in Table 2.1. In this table all cases are described by variables called *attributes* and one variable called a *decision*. The set of all attributes is denoted by A. The decision is denoted by d. The set of all cases is denoted by U. In Table 2.1 the attributes are *Length*, *Width*, *Height* and *Weight*, while the decision is *Quality*. Additionally, $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$. A *concept* is the set of all cases with the same decision value. In Table 2.1 there are three concepts, $\{1, 2, 3\}$, $\{4, 5\}$ and $\{6, 7, 8\}$.

Table 2.1 A numerical data set

Case	Attributes				Decision
	Length	Height	Width	Weight	Quality
1	4.7	1.8	1.7	1.7	High
2	4.5	1.4	1.8	0.9	High
3	4.7	1.8	1.9	1.3	High
4	4.5	1.8	1.7	1.3	Medium
5	4.3	1.6	1.9	1.7	Medium
6	4.3	1.6	1.7	1.3	Low
7	4.5	1.6	1.9	0.9	Low
8	4.5	1.4	1.8	1.3	Low

Let a be a numerical attribute, let p be the smallest value of a and let q be the largest value of a. During discretization, the domain $[p, q]$ of the attribute a is divided into the set of k intervals,

$$\{[a_{i_0}, a_{i_1}), [a_{i_1}, a_{i_2}), \ldots, [a_{i_{k-2}}, a_{i_{k-1}}), [a_{i_{k-1}}, a_{i_k}]\},$$

where $a_{i_0} = p$, $a_{i_k} = q$, and $a_{i_l} < a_{i_{l+1}}$ for $l = 0, 1, \ldots, k-1$. The numbers a_{i_1}, a_{i_2},..., $a_{i_{k-1}}$ are called *cut-points*. Such intervals are denoted by

$$a_{i_0} \ldots a_{i_1}, a_{i_1} \ldots a_{i_2}, \ldots, a_{i_{k-2}} \ldots a_{i_{k-1}}, a_{i_{k-1}} \ldots a_{i_k}.$$

For any nonempty subset B of the set A of all attributes, an *indiscernibility* relation $IND(B)$ is defined, for any $x, y \in U$, in the following way

$$(x, y) \in IND(B) \text{ if and only if } a(x) = a(y) \text{ for any } a \in B, \tag{2.1}$$

where $a(x)$ denotes the value of the attribute $a \in A$ for the case $x \in U$. The relation $IND(B)$ is an equivalence relation. The equivalence classes of *IND(B)* are denoted by $[x]_B$.

A partition on U is the set of all equivalence classes of $IND(B)$ and is denoted by B^*. Sets from $\{d\}^*$ are concepts. For example, for Table 2.1, if $B = \{Length\}$, $B^* = \{\{1, 3\}, \{2, 4, 7, 8\}, \{5, 6\}\}$ and $\{d\}^* = \{\{1, 2, 3\}, \{4, 5\}, \{6, 7, 8\}\}$. A data set is consistent if $A^* \le \{d\}^*$, i.e., if for each set X from A^* there exists set Y from $\{d\}^*$ such that $X \subseteq Y$. For the data set from Table 2.1, each set from A^* is a singleton, so this data set is consistent.

We quote the Dominant Attribute discretization algorithm [12, 14]. The first task is sorting of the attribute domain. Potential cut-points are selected as means of two consecutive numbers from the sorted attribute domain. For example, for *Length* there are two potential cut-points: 4.4 and 4.6.

Let S be a subset of the set U. An entropy $H_S(a)$ of an attribute a, with the values $a_1, a_2, \ldots, a_n$ is defined as follows

$$-\sum_{i=1}^{n} p(a_i) \cdot \log p(a_i), \tag{2.2}$$

where $p(a_i)$ is a probability (relative frequency) of the value a_i of the attribute a, $a_1, a_2, \ldots, a_n$ are all values of a in the set S, logarithms are binary, and $i = 1, 2, \ldots, n$.

A conditional entropy for the decision d given an attribute a, denoted by $H_S(d|a)$ is

$$\sum_{i=1}^{n} p(a_i) \cdot \sum_{j=1}^{m} p(d_j|a_i) \cdot log\ p(d_j|a_i), \tag{2.3}$$

$p(d_j|a_i)$ is the conditional probability of the value d_j of the decision d given the value a_i of a and $d_1, d_2, \ldots, d_m$ are all values of d in the set S. The main ideas of Dominant Attribute discretization algorithm are presented below.

Procedure Dominant Attribute

Input: a set U of cases, a set A of attributes, a set $\{d\}^*$ of concepts
Output: a discretized data set
$\{A\}^* := \{U\}$;
$\{B\}^* := \emptyset$;
while $\{A\}^* \not\leq \{d\}^*$ **do**
    X := SelectBlock($\{A\}^*$);
    a := BestAttribute(X);
    c := BestCutPoint(X, a);
    $\{S_1, S_2\}$:= Split(X, c);
    $\{B\}^* := \{B\}^* \cup \{S_1, S_2\}$;
    $\{A\}^* := \{B\}^*$;
end

In the Dominant Attribute discretization method, initially we need to select the dominant attribute, defined as an attribute with the smallest entropy $H_S(a)$. The process of computing of $H_U(Length)$ is illustrated in Fig. 2.1. In the Figs. 2.1 and 2.2 l stands for low, m for medium, and h for high, where {low, medium, high} is the domain of *Quality*.

$$H_U(Length) = \frac{1}{4}\left(-\frac{1}{2} \cdot \log \frac{1}{2}\right)2 + \frac{1}{2}\left(\left(-\frac{1}{4} \cdot \log \frac{1}{4}\right)2 - \frac{1}{2} \cdot \log \frac{1}{2}\right) + \frac{1}{4} \cdot 0 = 1.$$

Similarly, we compute remaining three conditional entropies: $H_U(Height) \approx 0.940$, $H_U(Width) \approx 1.439$ and $H_U(Weight) = 1.25$. We select *Height* since its entropy is the smallest.

Let a be an attribute and q be a cut-point of the attribute a that splits the set S into two subsets, S_1 and S_2. The conditional entropy $H_S(d|a, q)$ is defined as follows

$$\frac{|S_1|}{|S|} H_{S_1}(d|a) + \frac{|S_2|}{|S|} H_{S_2}(d|a), \tag{2.4}$$

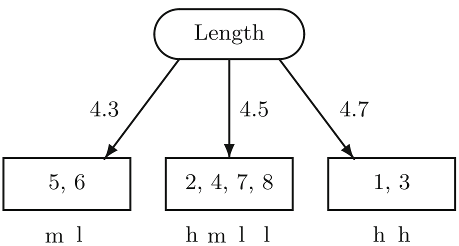

Fig. 2.1 Computing conditional entropy $H_U(Length)$

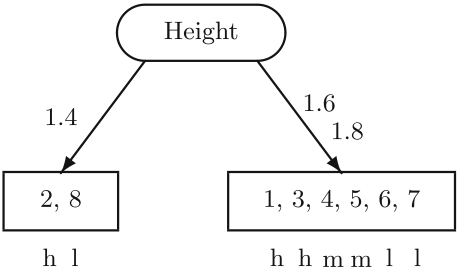

Fig. 2.2 Computing conditional entropy $H_U(Quality|Height, 1.5)$

where $|X|$ denotes the cardinality of the set X. The cut-point q for which the conditional entropy $H_S(d|a, q)$ has the smallest value is selected as the best cut-point.

Thus, the next step is to find the best cutpoint for *Height*. There are two candidates: 1.5 and 1.7. We need to compute two conditional entropies, namely $H_U(Quality|Height, 1.5)$ and $H_U(Quality|Height, 1.7)$. Computing the former entropy is illustrated in Fig. 2.2. $H_U(Quality|Height, 1.5) = \frac{1}{4}(-\frac{1}{2} \cdot \log \frac{1}{2})2 + \frac{3}{4}(-\frac{1}{3} \cdot \log \frac{1}{3})3 \approx 1.439$.

Similarly, $H_U(Quality|Height, 1.7) \approx 1.201$. We select the cut-point with the smaller entropy, i.e., 1.7.

After any selection of a new cut-point we test whether discretization is completed, i.e., if the data set with discretized attributes is consistent. So far, we discretized only one attribute, *Height*. Remaining, not yet discretized attributes, have values $p..q$, where p is the smallest attribute value and q is the largest attribute value. The corresponding table is presented in Table 2.4. For Table 2.4, $A^* = \{\{1, 3, 4\}, \{2, 5, 6, 7, 8\}\}$, so $A^* \not\leq \{d\}^*$, the data set from Table 2.4 needs more discretization. The cut-point 1.7 of *Height* splits the original data set from Table 2.1 into two smaller subtables, the former with the cases 1, 3 and 4 and the latter with the cases 2, 5, 6, 7 and 8. The former subtable is presented as Table 2.2, the latter as Table 2.3. The remaining computing is conducted by recursion. We find the best attribute for Table 2.2, then the best cut-point, and we check whether the currently dicretized data set is consistent. If not, we find the best attribute for Table 2.3, the best cut-point, and we check again whether the currently dicretized data set is consistent. If not, we compute new numerical data sets that result from current cut-points, and again, compute the best attribute, the best cut-point, and check whether the currently discretized data set is consistent.

Table 2.2 Numerical data set restricted to $\{1, 3, 4\}$

	Attributes				Decision
Case	Length	Height	Width	Weight	Quality
1	4.6..4.7	1.7..1.8	1.7..1.9	1.5..1.7	High
3	4.6..4.7	1.7..1.8	1.7..1.9	0.9..1.5	High
4	4.3..4.6	1.7..1.8	1.7..1.9	0.9..1.5	Medium

Table 2.3 A numerical data set restricted to {2, 5, 6, 7, 8}

	Attributes				Decision
Case	Length	Height	Width	Weight	Quality
2	4.3..4.6	1.4..1.7	1.7..1.9	0.9..1.5	High
5	4.3..4.6	1.4..1.7	1.7..1.9	1.5..1.7	Medium
6	4.3..4.6	1.4..1.7	1.7..1.9	0.9..1.5	Low
7	4.3..4.6	1.4..1.7	1.7..1.9	0.9..1.5	Low
8	4.3..4.6	1.4..1.7	1.7..1.9	0.9..1.5	Low

Table 2.4 Numerical data set with discretized *Height*

	Attributes				Decision
Case	Length	Height	Width	Weight	Quality
1	4.3..4.7	1.7..1.8	1.7..1.9	0.9..1.7	High
2	4.3..4.7	1.4..1.7	1.7..1.9	0.9..1.7	High
3	4.3..4.7	1.7..1.8	1.7..1.9	0.9..1.7	High
4	4.3..4.7	1.7..1.8	1.7..1.9	0.9..1.7	Medium
5	4.3..4.7	1.4..1.7	1.7..1.9	0.9..1.7	Medium
6	4.3..4.7	1.4..1.7	1.7..1.9	0.9..1.7	Low
7	4.3..4.7	1.4..1.7	1.7..1.9	0.9..1.7	Low
8	4.3..4.7	1.4..1.7	1.7..1.9	0.9..1.7	Low

Table 2.5 Completely discretized data set

	Attributes				Decision
Case	Length	Height	Width	Weight	Quality
1	4.6..4.7	1.7..1.8	1.7..1.9	1.5..1.7	High
2	4.3..4.6	1.4..1.5	1.7..1.9	0.9..1.1	High
3	4.6..4.7	1.7..1.8	1.7..1.9	1.1..1.5	High
4	4.3..4.6	1.7..1.8	1.7..1.9	1.1..1.5	Medium
5	4.3..4.6	1.5..1.7	1.7..1.9	1.5..1.7	Medium
6	4.3..4.6	1.5..1.7	1.7..1.9	1.1..1.5	Low
7	4.3..4.6	1.5..1.7	1.7..1.9	0.9..1.1	Low
8	4.3..4.6	1.4..1.5	1.7..1.9	1.1..1.5	Low

Our finally discretized data set, presented in Table 2.5, is consistent. The last step is an attempt to merge successive intervals. Such attempt is successful if a new discretized data set is still consistent. It is not difficult to see that all cut-points are necessary. For example, if we remove cut-point 4.6 for *Length*, cases 3 and 4 will be indistinguishable, while these two cases belong to different concept.

Note that the discretized data set, presented in Table 2.5, has four attributes and five cut-points. One of attributes, *Width*, is redundant. Thus, the reduced attribute set consists of three attributes: *Length*, *Height* and *Weight*.

2.3 Multiple Scanning Discretization

The Multiple Scanning discretization is also based on conditional entropy. However, this method uses a different strategy. The entire attribute set is scanned t times, where t is a parameter called the total number of scans. During every scan the best cut-point is computed for all attributes. The parameter t is provided by the user. If t is too small, the discretized data set is not consistent and Dominant Attribute method is used.

Procedure Multiple Scanning

```
Input: a set U of cases, a set A of attributes, a set {d}* of concepts, a number of scans t
Output: a discretized data set
{A}* := {U};
foreach scan := 1 to t do
    if {A}* ≤ {d}* then
        break;
    end
    C := ∅;
    foreach a ∈ A do
        cut_point := BestCutPointMS({A}*, a);
        C := C ∪ {cut_point};
    end
    {A}* := Split({A}*, C);
end
```

The main ideas of the Multiple Scanning algorithm are presented above. For details see [10, 11, 13, 14]. Obviously, for the same data set, data sets discretized by the Dominant Attribute and Multiple Scanning methods are, in general, different. We consider the Multiple Scanning method as auxiliary one.

Since during every scan all attributes are discretized, usually the discretized data set has all original attributes. The only chance to eliminate some attributes is during the last step, i.e., merging successive intervals.

2.4 Experiments

Our experiments were conducted on 15 data sets with numerical attributes presented in Table 2.6. All of these data sets are accessible at the University of California at Irvine *Machine Learning Repository*. First we discretized all data sets using Dominant Attribute method. Then we identified data sets with single interval attributes, i.e., discretized values in which the entire domain of a numerical attribute was mapped into a single interval. For two data sets, *Abalone* and *Iris*, no single interval attributes were discovered, so these two data sets were removed from further experiments. For

Table 2.6 Data sets

Data set	Cases	Number of attributes	Concepts
Abalone	4,177	8	29
Australian	690	14	2
Bankruptcy	66	5	2
Bupa	345	6	2
Connectionist Bench	208	60	2
Echocardiogram	74	7	2
E coli	336	8	8
Glass	214	9	6
Image Segmentation	210	19	7
Ionoshere	351	34	2
Iris	150	4	3
Pima	768	8	2
Wave	512	21	3
Wine Recognition	178	13	3
Yeast	1,484	8	9

any data set with single interval attributes, a new data set with numerical attributes was created by removing single interval attributes from the original, numerical data set. Such data sets are called *reduced*.

Reduced data sets are presented in Table 2.7. As it was observed in Sect. 2.3, Multiple Scanning discretization seldom produces reduced data sets. In our experiments, Multiple Scanning produced reduced data sets only for three original data sets: *Connectionist Bench, Image Segmentation* and *Ionosphere*, so during analysis of experimental results Multiple Scanning was ignored.

All numerical data sets, original and reduced, were subjected to single ten-fold cross-validation using C4.5 system [18]. The system C4.5 was selected as well-known classifier. Note that C4.5 has an internal discretization method similar to Dominant Discretization algorithm. Error rates computed by C4.5 and ten-fold cross-validation are presented in Table 2.8. For our results we used the Wilcoxon matched-pairs two-tailed test with 5% significance level. We conclude that the error rate is significantly larger for reduced data sets. An additional argument for better quality of original data sets was reported in [10, 11, 13, 14]. Multiple Scanning discretization method was better than other discretization methods since in the majority of data sets discretized by Multiple Scanning all discretized attributes have more than one interval.

Table 2.7 Reduced data sets

Data set	Number of single-interval attributes for data sets reduced by	
	Dominant Attribute	Multiple Scanning
Abalone	0	0
Australian	9	0
Bankruptcy	3	0
Bupa	2	0
Connectionist Bench	57	8
Echocardiogram	3	0
E coli	3	0
Glass	3	0
Image Segmentation	15	1
Ionoshere	29	1
Iris	0	0
Pima	2	0
Wave	15	0
Wine Recognition	9	0
Yeast	3	0

Table 2.8 C4.5 error rate, data sets reduced by dominant attribute

Name	Original data set	Reduced data set
Australian	16.09	15.36
Bankruptcy	6.06	12.12
Bupa	35.36	35.94
Connectionist Bench	25.96	25.96
Echocardiogram	28.38	44.59
E coli	17.86	19.35
Glass	33.18	33.18
Image Segmentation	12.38	10.48
Ionoshere	10.54	11.97
Pima	25.13	26.95
Wave	26.37	32.42
Wine Recognition	8.99	8.99
Yeast	44.41	48.45

We compared complexity of decision trees generated by C4.5 from original and reduced data sets as well. Results are presented in Tables 2.9 and 2.10. The tree depth is the number of edges on the longest path between the root and any leaf. The tree size is the total number of nodes of the tree. These numbers are reported by the C4.5

Table 2.9 C4.5 tree depth, data sets reduced by dominant attribute

Name	Original data set	Reduced data set
Australian	11	5
Bankruptcy	1	3
Bupa	8	8
Connectionist Bench	7	2
Echocardiogram	4	3
E coli	9	7
Glass	8	8
Image Segmentation	8	7
Ionoshere	11	10
Pima	9	7
Wave	10	9
Wine Recognition	3	3
Yeast	22	20

Table 2.10 C4.5 tree size, data sets reduced by dominant attribute

Name	Original data set	Reduced data set
Australian	63	11
Bankruptcy	3	7
Bupa	51	33
Connectionist Bench	35	5
Echocardiogram	9	7
E coli	43	37
Glass	45	45
Image Segmentation	25	25
Ionoshere	35	25
Pima	43	35
Wave	85	63
Wine Recognition	9	9
Yeast	371	411

system. Using the same Wilcoxon test we conclude that decision trees generated from reduced trees are simpler. The depth of decision trees is smaller for reduced data sets with significance level 5%. On the other hand, the size of decision trees is smaller with significance level 10%.

2.5 Conclusions

Let us recall that our main objective is to compare quality of numerical data sets, original and reduced, using the C4.5 decision tree generation system. Our experiments on 15 numerical data sets show that for a Dominant Attribute discretization method the error rate computed by C4.5 and ten-fold cross-validation is significantly larger (5% significance level, two-tailed test) for the reduced data sets than for the original data sets. However, decision trees generated from the reduced data sets are significantly simpler than the decision trees generated from the original data sets. Thus, if our top priority is accuracy, the original data sets should be used. On the other hand, if all what we want is simplicity we should use reduced data sets.

References

1. Blajdo, P., Grzymala-Busse, J.W., Hippe, Z.S., Knap, M., Mroczek, T. L., Piatek : A comparison of six approaches to discretization—a rough set perspective. In: Proceedings of the Rough Sets and Knowledge Technology Conference, pp. 31–38 (2008)
2. Bruni, R., Bianchi, G.: Effective classification using a small training set based on discretization and statistical analysis. IEEE Trans. Knowl. Data Eng. **27**(9), 2349–2361 (2015)
3. Chmielewski, M.R., Grzymala-Busse, J.W.: Global discretization of continuous attributes as preprocessing for machine learning. Int. J. Approx. Reason. **15**(4), 319–331 (1996)
4. Clarke, E.J., Barton, B.A.: Entropy and MDL discretization of continuous variables for bayesian belief networks. Int. J. Intell. Syst. **15**, 61–92 (2000)
5. de Sa, C.R., Soares, C., Knobbe, A.: Entropy-based discretization methods for ranking data. Inf. Sci. **329**, 921–936 (2016)
6. Elomaa, T., Rousu, J.: General and efficient multisplitting of numerical attributes. Mach. Learn. **36**, 201–244 (1999)
7. Fayyad, U.M., Irani, K.B.: Multi-interval discretization of continuous-valued attributes for classification learning. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence, pp. 1022–1027 (1993)
8. Fayyad, U.M., Irani, K.B.: On the handling of continuous-valued attributes in decision tree generation. Mach. Learn. **8**, 87–102 (1992)
9. Garcia, S., Luengo, J., Sáez, J.A., Lopez, V., Herrera, F.: A survey of discretization techniques: taxonomy and empirical analysis in supervised learning. IEEE Trans. Knowl. Data Eng. **25**(4), 734–750 (2013)
10. Grzymala-Busse, J.W., Mroczek, T.: A comparison of two approaches to discretization: multiple scanning and C4.5. In: Proceedings of the 6-th International Conference on Pattern Recognition and Machine Learning, pp. 44–53 (2015)
11. Grzymala-Busse, J.W.: A multiple scanning strategy for entropy based discretization. In: Proceedings of the 18th International Symposium on Methodologies for Intelligent Systems, pp. 25–34 (2009)
12. Grzymala-Busse, J.W.: Discretization of numerical attributes. In: Kloesgen, W., Zytkow, J. (eds.) Handbook of Data Mining and Knowledge Discovery, pp. 218–225. Oxford University Press, New York, NY (2002)
13. Grzymala-Busse, J.W.: Discretization based on entropy and multiple scanning. Entropy **15**, 1486–1502 (2013)
14. Grzymala-Busse, J.W., Mroczek, T.: A comparison of four approaches to discretization based on entropy. Entropy **18**, 1–11 (2016)

15. Jiang, F., Sui, Y.: A novel approach for discretization of continuous attributes in rough set theory. Knowl. Based Syst. **73**, 324–334 (2015)
16. Kohavi, R., Sahami, M.: Error-based and entropy-based discretization of continuous features. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 114–119 (1996)
17. Nguyen, H.S., Nguyen, S.H.: Discretization methods in data mining. In: Polkowski, L., Skowron, A. (eds.) Rough Sets in Knowledge Discovery 1: Methodology and Applications, pp. 451–482. Physica-Verlag, Heidelberg (1998)
18. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo (1993)
19. Rahman, M.D., Islam, M.Z.: Discretization of continuous attributes through low frequency numerical values and attribute interdependency. Expert Syst. Appl. **45**, 410–423 (2016)
20. Sang, Y., Qi, H., Li, K., Jin, Y., Yan, D., Gao, S.: An effective discretization method for disposing high-dimensional data. Inf. Sci. **270**, 73–91 (2014)
21. Stefanowski, J.: Handling continuous attributes in discovery of strong decision 0 rules. In: Proceedings of the First Conference on Rough Sets and Current Trends in Computing, pp. 394–401 (1998)
22. Stefanowski, J.: Algorithms of Decision Rule Induction in Data Mining. Poznań University of Technology Press, Poznań, Poland (2001)

http://www.springer.com/978-3-319-67587-9

Advances in Feature Selection for Data and Pattern
Recognition
Stanczyk, U.; Zielosko, B.; Jain, L.C. (Eds.)
2018, XVIII, 328 p. 37 illus., 20 illus. in color., Hardcover
ISBN: 978-3-319-67587-9