Preserving Informative Content of Condition Attributes in Data Transformations for CRSA

Urszula Stańczyk[1][0000−0002−5071−7187], Beata Zielosko[2][0000−0003−3788−1094], and Grzegorz Baron[1][0000−0001−8613−631X]

[1] Department of Computer Graphics, Vision and Digital Systems, Silesian University of Technology, Akademicka 2A, 44-100 Gliwice, Poland
{urszula.stanczyk,grzegorz.baron}@polsl.pl
[2] Institute of Computer Science, University of Silesia in Katowice, Będzińska 39, 41-200 Sosnowiec, Poland
beata.zielosko@us.edu.pl

Abstract. The research work described in the paper addressed the preparation of the input data for the classical rough set approach, with the aim of preserving the informative content of all condition attributes. Instead of ignoring attributes whose values are assigned to a single interval by a supervised discretisation algorithm, such attributes are subjected to unsupervised discretisation processing. In order to examine the informativeness of attributes undergoing the fusion of discretisation methods, reducts and decision rules were induced as popular forms of knowledge representation, especially in the framework of rough set theory. The results obtained were studied from the point of view of the characteristics of knowledge representations and the performance of rule-based classifiers evaluated with test sets discretised in different ways. The conducted experiments demonstrate the validity of the investigated approach.

Keywords: CRSA · Decision Reduct · Decision Rule · Discretisation · Decision-Making.

1 Introduction

Attributes play a key role in decision-making because they provide the necessary information based on which decisions are made [10]. The influence of features and their types can be analysed in the context of attribute selection and ranking construction. This translates into machine learning processes and the discovery of patterns in the data [17]. Depending on the application needed, various methods of attribute transformation exist, for example, scaling values to fit a certain range, encoding categorical values referred to as one-hot encoding, and feature extraction for dimensionality reduction [23]. The objective of the research presented was to analyse the impact of attribute transformation as a result of discretisation, aimed at characteristics of knowledge patterns present in the data.

Transformation of continuous attribute values into discrete or nominal ones with a finite number of intervals can be performed in many ways [7]. If a discretiser takes into account class information to find proper intervals among ranges

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

of attribute values, it is called supervised. If class information is omitted during the discretisation and the number of intervals is provided as an input parameter, it is called unsupervised. An important issue during the discretisation process is the preservation of attributes' information value. It has implications for the patterns discovered in the data and the effectiveness of the prediction models. Discretisation, although it can simplify the data and remove possible noise from them, can also lead to irrecoverable loss of important information.

Given that both supervised and unsupervised methods have their shortcomings, in the work an approach involving a fusion of discretisation methods is investigated, with fusion understood as a merging of distinctively different approaches applied to parts of the data to return an entire discretised dataset. In the case of attributes that are given 1-bin values as a result of a supervised discretisation process, that is, values that do not introduce any variation in the attribute domain, an unsupervised discretisation method is applied. In consequence, all available features have meaningful representations in a discrete domain. The proposed attribute transformation process is implemented in two modes: (i) as an independent transformation when the constructed intervals are created independently in the training and test sets, and (ii) as a dependent transformation when the interval definitions found for the training sets are used to find a discrete representation for the test sets.

Keeping the informative content of the attributes is all important for data exploration algorithms that require categorical variables, such as the methods used in the framework of rough sets theory [14]. The notions of reducts and decision rules lead to discovering patterns in the data, supporting transparency and interpretability of results. The characteristics of these patterns, such as cardinality of reducts or length of rules, can be influenced by the type and mode of discretisation process to which the input data are subjected.

Extensive experiments were carried out, with exhaustive search algorithms employed for induction of reducts and rules from stylometric data [13] discretised in various ways. The results confirm that the fusion of supervised and unsupervised discretisation methods, aimed at ensuring the interpretability of the data and maintaining its distinguishability, contributed to information preservation and pattern discovery from the data. Both considered modes of the discretisation procedures, dependent and independent, led to cases of increased performance of the rule classifiers, validating the presented research framework.

The paper consists of four main sections. The introduction is followed by Section 2 related to the fundamentals of rough sets theory and popular tools for knowledge representation. Section 3 contains information on properties of discovered knowledge patterns, including the performance of rule-based classifiers. Section 4 contains conclusions and comments on future works.

2 Knowledge Patterns Mined by Rough Set Approach

The rough set theory allows data to be analysed and knowledge to be extracted from them, based on information granules. These are collections of objects, in-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

distinguishable from the point of view of the available knowledge. In the rough sets perspective, instead of individual objects, granules of indiscernible objects are used. Imprecise concepts are approximated, replaced by a pair of precise concepts called the lower and upper approximation of the rough concept [15]. Therefore, the indiscernibility relation $IND(B)$ plays an important role.

Let $S = (U, A)$ be an information system where U is a non-empty, finite set of objects and $A = \{a_1, \ldots, a_n\}$ is a non-empty, finite set of attributes. $a_i : U \to V_{a_i}$, where V_{a_i} is the set of values of attribute a_i called the domain of a_i. Indiscernibility relation $IND(B)$ is defined as follows:

$$IND(B) = \{(x, y) \in U \times U : \forall_{a \in B} \quad a(x) = a(y)\}. \tag{1}$$

Its equivalence classes are defined by $[x]_{IND(B)} = \{y \in U : (x, y) \in IND(B)\}$, for any object $x \in U$.

Let $B \subseteq A$ and $X \subseteq U$. Then the lower $\underline{B}(X)$ and upper $\overline{B}(X)$ approximations of X with respect to B are defined respectively:

$$\underline{B}(X) = \{x \in U : [x]_{IND(B)} \subseteq X\} \tag{2}$$

$$\overline{B}(X) = \{x \in U : [x]_{IND(B)} \cap X \neq \emptyset\}. \tag{3}$$

The set $BN_B(X) = \overline{B}(X) \setminus \underline{B}(X)$ is called a boundary region of the concept X. It consists of all elements that cannot be classified to the set X or its complement, by employing available knowledge. So, rough set theory expresses imprecision by employing a boundary region of a set.

Patterns can be represented in various ways. A decision table, used for tabular data presentation, is the simplest form. These structures allow the application of methods for the induction of decision rules and reducts, which are also recognised as popular forms of discovered patterns mined by the rough sets approach.

2.1 Datasets and Decision Tables

Depending on the type of data, the context and the purpose of the analysis, the patterns in the data can be represented in different ways. In rough set theory, the most basic representation of the data is provided by a decision table. This form is widely used because it is simple and transparent and allows analysis, optimisation, and automation of decision-making processes. It can be treated as a special type of information system in which the decision attribute d is distinguished. Formally, a decision table is defined as: $T = (U, A, \{d\})$ [14], where U is a non-empty finite set of objects, $A = \{a_1, \ldots, a_n\}$ is a non-empty finite set of condition attributes, and d, $d \notin A$, is a distinguished attribute called a decision or class label, with values $V_d = \{d_1, \ldots, d_{|V_d|}\}$.

The data used in the research were represented in the form of decision tables related to the two prepared datasets. The datasets were dedicated to the task of authorship attribution in the domain of stylometric analysis of texts [13], treated as a classification problem. Recognition is based on authorial profiles built in the training phase, which are then measured against during testing. Profiles refer to

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

stylometric markers, such as the frequency of usage of function words, elements of the text structure, or other descriptors [22].

The authors studied were four renowned writers, Edith Wharton, Mary Johnston, Henry James, and Thomas Hardy. They were put in pairs: the female writer dataset (F-writers) based on selected works of Wharton and Johnston, and the male writer dataset (M-writers) based on some novels of James and Hardy. Each dataset included three sets (decision tables), one training set, and two test sets. Lexical markers were used as condition attributes, taking frequencies of usage for twelve 2-letter function words as follows: of, in, to, on, at, by, or, as, if, up, so, no. The continuous values of attributes were calculated for text samples obtained by partitioning longer texts into smaller chunks of comparable size. The single decision attribute provided through its nominal values the names of authors for samples. Each set (decision table) was prepared for binary classification with balanced classes.

2.2 Transformations of Attribute Domains

The classical rough set approach requires the attributes to be either nominal or discrete to infer knowledge from them. When the input domain is continuous as in the described research, attribute domains need to be transformed before they can be mined. Discretisation algorithms for each translated variable find representation through some finite number of intervals (bins). The procedure can take into account information on class labels assigned to samples in supervised proceedings or focus entirely on processed values in unsupervised approaches [7].

Fayyad and Irani (denoted dsF) [6], and Kononenko (denoted dsK) [12] belong to popular supervised discretisation methods. Both involve entropy for the assessment of cut-points of constructed intervals for discretised attributes and Minimal Description Length (MDL) principle [16] as a stopping criterion. In general, the discretisation process starts with a single bin containing all the values of the attribute to be discretised and then this interval is recursively subdivided into smaller ones until a stopping criterion is reached.

Supervised discretisation can be perceived as possessing characterisation property with respect to transformed attributes because, for each variable, the procedure returns a specific list of established intervals. For the two datasets used, these results are shown in Table 1, where it can be observed that for a significant part of the variables, only single bins were found. It is a consequence of the discretisation procedures that stopped once it was detected that forming further smaller intervals would not be supportive for class distinction. Accepting such a representation would mean disregarding completely informative content of these variables and effectively reducing feature sets. Instead, in the research, a different approach was adopted. All attributes, for which by supervised discretisation just one bin was found, were subjected to unsupervised transformations.

Equal width binning (denoted duw) belongs to the popular unsupervised discretisation algorithms. This method is simple, but it is also sensitive to the number of bins defined by the user. If the values of discretised attributes are

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

Table 1. Characteristics of attributes by supervised discretisation using the Fayyad and Irani (dsF) and Kononenko (dsK) approaches for training sets.

	F-writers		M-writers	
	dsF	dsK	dsF	dsK
Bins	Attributes			
1	if in no or so up	if in no or so up	as of no on so to up	to on as up so no
2	as at by	of at by as	at if or	of at or if
3	of on to	to on	by in	in by

unevenly distributed, some information can be lost after the discretisation process. In general, this method orders the values of a continuous attribute, finds the minimum and maximum values, and then divides the range into the k equal width discrete intervals, where k is the input parameter.

The input data explored with the classical rough set approach included 20 variants of each dataset: one from supervised discretisation by the Fayyad and Irani algorithm (dsF), one from supervised discretisation by the Kononenko procedure (dsK), and 18 variants for combined supervised-unsupervised processing, based on either Fayyad and Irani or Kononenko approach supported with equal width binning with varying the number of bins from 2 to 10 (from dsFuw02 to dsFuw10, and from dsKuw02 to dsKuw10). All variants were mined to infer knowledge patterns in the form of decision reducts and decision rules.

20 variants of a training set were accompanied by twice as many variants of each test set as they were transformed in two different modes [3]. In independent processing (denoted test independent, Tind), the constructed intervals disregarded all other sets and were based entirely on the characteristics of each discretised set. Another approach involved using definitions of intervals found for training sets to find a discrete representation for test sets (denoted test on learn, ToL).

2.3 Decision Reducts

Reducts are a popular feature selection method used in rough set theory [24]. They are minimal sets of attributes that provide the same classification of objects as the entire set of attributes. There are different types of reducts and algorithms for their construction.

Decision reduct [15] is a minimal set of attributes $B \subseteq A$ that determines the decision values within T. Formally, a subset of attributes $B \subseteq A$ is a decision reduct for T if and only if it is an irreducible subset of attributes such that $IND(B) \subseteq IND(d)$. For $RED(A)$ being a set of all reducts B_i of A, $B_i \in RED(A)$, $\forall i \in \{1, \ldots, |RED(A)|\}$, the intersection of reducts is called a core, $Core(A) = \bigcap_{i=1}^{|RED(A)|} B_i$. Core contains all indispensable attributes from A.

The number of attributes that form a reduct is important as it affects the complexity and interpretability of the data model, and the quality of the classification, by providing a minimal but sufficient set of attributes to describe the data. If an attribute appears in the core, it means that it appears in all reducts. Therefore, deletion of it would affect the loss of information.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

The problem of constructing reducts with the minimum number of attributes is NP-hard [23]. In addition, the number of all reducts that exist in a given decision table, consisting of m attributes, is equal to $N(m) = \binom{m}{m/2}$. These factors caused good conditions for the invention of approximate algorithms for calculating reducts, using heuristics based on genetic or greedy algorithms, attribute importance, mutual information, and many other methods [9].

Short reducts can be used for dimensionality reduction, feature selection, or building decision rule-based classifiers. In the worst case, computing all shortest reducts has the same exponential complexity as computing all reducts. However, in practice, computing only the shortest reducts is usually much faster than computing the entire set of reducts [8]. Therefore, in practical applications, computing the former may be preferred to the latter.

Induced decision reducts can be used to characterise the importance of particular attributes and show how it is reflected by including them in reducts of varying quality [17, 18]. In this work, reducts are important elements of the proposed research methodology related to the discovery of changing knowledge patterns from stylometric data based on the fusion of discretisation methods.

2.4 Decision Rules

Decision rules are a popular form of representation of patterns present in the data. *if then* rules are the best known form. The part *if* is the conditional part of the rule, which consists of descriptors (pairs $a_i = v_{i_j}$), and the part *then* is defined as the decision part. In general, the rule can be written as:

$$(a_{i_1} = v_1) \wedge \ldots \wedge (a_{i_m} = c_m) \rightarrow d, \tag{4}$$

where $a_{i_1}, \ldots, a_{i_m} \in A$, $v_1, \ldots, v_m$ are values of the attribute a_i, and d is a decision (class label). Therefore, a rule explicitly lists the conditions and their values that lead to a particular decision.

Again, the problem of constructing decision rules of minimal length is NP-hard [23], although there are some algorithms for the construction of sufficiently short rules. In general, many approaches to inducing decision rules exist. Popular ones include: methods based on rough sets theory [15], rule induction from decision trees [25], approaches relaying on probabilistic methods [5], genetic and ant colony optimisation algorithms [11], or rule extraction from neural networks [1].

In this study, to induce decision rules, the exhaustive algorithm implemented in the Rough Set Exploration System (RSES) was used [4]. It allows finding all rules optimal in terms of length. This measure plays an important role in understanding and interpreting the patterns in the data. For all data variants resulting from merged supervised and unsupervised discretisation approaches, decision rules were induced, which returned rule sets with varying characteristics.

3 Properties of Induced Knowledge Patterns

The number of attributes that make up a reduct can be seen as its evaluation measure. From the point of view of induced knowledge, shorter reducts offer

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

a higher reduction in dimensionality. They may be preferred because they are easier to understand and interpret.

For decision rules, the length is equal to the number of descriptors in the premise part of the rule, and the support is the number of objects matched to the rule. Shorter rules help generalisation, while high support allows the discovery of essential patterns in the data. Longer rules increase the risk of over-fitting in the classification process, and very low support values call attention to so-called outliers. Due to their importance, these characteristics of the knowledge patterns discovered can be used as quality measures [19, 21].

The induction of short rules and reducts coincides with the MDL principle [16]. The cardinality of the reducts as well as the length of the decision rules and their support affect the comprehensibility, efficiency, and generality of the model. The characteristic features of the input domain and their transformations have an impact on these induced patterns. When the perspective is reversed, reducts and rules can be seen as characterising the attributes and their roles in decision-making, indicate their relevance.

3.1 Attributes and Quality of Decision Reducts

Decision reducts were induced with the exhaustive algorithm implemented in the RSES system from all investigated data variants. The characteristics of the reducts found are included in Table 2. It can be observed that with the higher numbers of intervals constructed for additionally transformed variables, the numbers of induced reducts rose as well. The trend, though not strictly monotonic, is clearly visible. Compared to the reduct sets obtained for data variants resulting from supervised discretisation, with supporting transformation from unsupervised processing, the average reduct length firstly steeply increased, to gradually decrease to values close to those observed at the beginning.

The characteristics of the reduct sets should be analysed while keeping in mind the number of conditional attributes considered. For the dsF and dsK data variants, it is significantly smaller, equal to half or even less than half of the number available when all attributes are represented meaningfully in a discrete space, which happens for a combination of supervised-unsupervised processing. For data discretised by a supervised algorithm, only single reducts are found, and all attributes with multiple bins constructed are included in these reducts. On the other hand, for extended and combined discretisation procedures, the attributes occur in reducts with varying frequency, which is presented in Table 3.

The table shows relations between the number of intervals constructed for a condition attribute and the number of times this attribute was included in decision reducts induced for all discrete data variants studied. The left part of the table (attributes separated with the vertical double line) contains the features which in supervised discretisation were assigned several bins for representation. The right part includes variables with a single interval representing from this processing, further divided into varying numbers of bins by unsupervised transformations, which is why this number is given as changing from 1 to 10.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

Table 2. Characteristics of sets of decision reducts inferred from the input data.

Data variant	F-writers				M-writers			
	Nr of reducts	Avg length	Min length	Max length	Nr of reducts	Avg length	Min length	Max length
dsF	1	6.0	6	6	1	5.0	5	5
dsFuw02	4	7.8	7	8	2	10.5	10	11
dsFuw03	33	7.7	7	10	4	9.8	9	10
dsFuw04	58	7.2	6	9	35	8.6	8	9
dsFuw05	97	7.0	6	8	84	7.6	6	9
dsFuw06	143	6.4	5	8	138	7.0	6	8
dsFuw07	115	6.4	5	8	165	6.7	6	8
dsFuw08	150	6.1	5	8	110	6.3	5	7
dsFuw09	168	6.0	5	7	157	6.1	5	8
dsFuw10	137	5.8	5	7	163	6.1	5	7
dsK	1	6.0	6	6	1	6.0	6	6
dsKuw02	6	8.7	8	9	3	11.0	11	11
dsKuw03	26	7.7	7	10	4	9.8	9	10
dsKuw04	58	7.3	6	9	20	8.4	8	9
dsKuw05	95	7.2	6	8	62	7.5	6	9
dsKuw06	145	6.6	5	8	87	6.9	6	8
dsKuw07	114	6.5	6	8	115	7.0	6	8
dsKuw08	151	6.2	5	8	82	6.4	5	8
dsKuw09	165	6.1	5	7	116	6.4	5	8
dsKuw10	124	5.8	5	7	145	6.2	5	7

The attributes presented in the two parts of the table were also compared with each other in terms of the number of cases in which these variables were included in the induced decision reducts. It is clear that the ones on the left-hand side of the table, supposedly more important and sufficient for class distinction, were not always predominantly used for reduct construction. In particular, for higher numbers of bins defined for attributes discretised by the unsupervised approach, these features were often more frequently included in reducts.

If a core of reducts is not an empty set, then it includes indispensable attributes. For the investigated reduct sets, in most cases the core was empty, but when it contained some features (denoted with bold font) they came both from the left and right part of the table, or even were found only in the group of features after unsupervised discretisation. This observation shows the impact of the discretisation process on the patterns present in the data and consequences of information fusion resulting from combined supervised with unsupervised processing of variables and their domains.

3.2 Sets of Decision Rules and Performance

Induced decision rules capture patterns discovered in the data. They can be considered and evaluated individually, but generally are rather analysed and used as sets (or lists). Within each set, some overall characteristics, in relation to the number of rules and the length and support of the rules, were included in Table 4 for both the female and male writer datasets in all their variants explored, differing in the discretisation approach applied to the condition attributes. Among other listed elements, minimal rule length and minimal rule support are not included, because for all data variants these two numbers had a constant value equal to 1.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

Table 3. Characteristics of attributes by sets of decision reducts inferred. When the core was non-empty, the attributes belonging to it are marked with bold font.

F-writers														
		Number of bins for attributes												
Data	Nr of	3			2			1...10						
variant	reducts	of	on	to	as	at	by	if	in	no	or	so	up	
dsF	1	1	1	1	1	1	1	0	0	0	0	0	0	
dsFuw02	4	**4**	**4**	**4**	**4**	0	2	2	3	**4**	0	2	2	
dsFuw03	33	25	31	**33**	21	5	24	18	21	21	19	17	20	
dsFuw04	58	34	53	38	37	9	33	36	33	48	31	28	38	
dsFuw05	97	59	65	58	52	23	52	66	65	71	63	52	53	
dsFuw06	143	78	92	80	83	26	62	77	91	86	85	79	80	
dsFuw07	115	66	76	61	60	22	56	56	69	76	78	63	54	
dsFuw08	150	79	84	71	72	29	64	76	94	94	89	84	75	
dsFuw09	168	77	102	91	90	21	69	95	86	95	94	97	87	
dsFuw10	137	71	76	61	57	17	72	79	73	76	70	72	72	
		Number of bins for attributes												
Data	Nr of	3		2				1...10						
variant	reducts	on	to	as	at	by	of	if	in	no	or	so	up	
dsK	1	1	1	1	1	1	1	0	0	0	0	0	0	
dsKuw02	6	**6**	**6**	**6**	4	4	**6**	3	4	**6**	0	4	3	
dsKuw03	26	24	**26**	21	4	18	18	14	15	18	13	13	16	
dsKuw04	58	54	38	39	13	33	34	36	32	49	31	29	38	
dsKuw05	95	63	65	53	31	50	57	66	62	73	60	55	52	
dsKuw06	145	94	80	83	35	70	80	79	91	90	92	82	79	
dsKuw07	114	77	61	60	27	54	65	60	65	81	78	61	55	
dsKuw08	151	86	75	75	36	67	80	73	94	91	90	89	82	
dsKuw09	165	103	95	89	21	71	74	90	82	99	93	100	85	
dsKuw10	124	77	62	52	13	60	58	67	67	67	65	74	62	

M-writers														
		Number of bins for attributes												
Data	Nr of	3		2			1...10							
variant	reducts	by	in	at	if	or	of	on	to	as	no	so	up	
dsF	1	1	1	1	1	1	0	0	0	0	0	0	0	
dsFuw02	2	**2**	**2**	**2**	**2**	**2**	1	**2**	1	**2**	1	**2**	**2**	
dsFuw03	4	**4**	2	**4**	**4**	**4**	2	**4**	2	3	2	**4**	**4**	
dsFuw04	35	23	30	24	22	25	21	21	22	**35**	24	30	25	
dsFuw05	84	53	40	42	47	51	52	54	47	53	56	58	84	
dsFuw06	138	86	55	83	63	71	71	79	88	83	98	102	92	
dsFuw07	165	99	85	87	81	80	96	95	85	93	101	110	100	
dsFuw08	110	56	49	42	54	38	61	50	48	**110**	58	59	69	
dsFuw09	157	95	63	72	65	60	96	73	88	84	84	89	90	
dsFuw10	163	76	62	82	73	80	77	84	85	113	73	81	103	
		Number of bins for attributes												
Data	Nr of	3		2				1...10						
variant	reducts	by	in	at	if	or	of	on	to	as	no	so	up	
dsK	1	1	1	1	1	1	1	0	0	0	0	0	0	
dsKuw02	3	**3**	**3**	**3**	2	**3**	**3**	**3**	2	**3**	**3**	**3**	2	
dsKuw03	4	**4**	2	**4**	**4**	**4**	2	**4**	2	3	2	**4**	**4**	
dsKuw04	20	13	20	11	11	16	6	14	13	**20**	13	20	12	
dsKuw05	62	44	31	30	35	36	30	44	37	38	38	48	62	
dsKuw06	87	49	34	49	41	38	20	59	61	56	66	67	63	
dsKuw07	115	74	86	58	50	56	46	66	72	71	71	78	73	
dsKuw08	82	45	33	30	40	28	33	43	36	**82**	49	49	57	
dsKuw09	116	61	49	63	49	46	55	72	68	73	66	73	69	
dsKuw10	145	76	61	70	66	63	59	86	81	107	69	74	94	

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

Table 4. Characteristics of sets of decision rules inferred from the input data.

Data variant	F-writers					M-writers				
	Nr of rules	Avg length	Max length	Avg support	Max support	Nr of rules	Avg length	Max length	Avg support	Max support
dsF	45	3.1	5	14.8	50	35	2.5	5	11.3	36
dsFuw02	663	4.0	7	5.8	53	1062	4.6	8	4.4	38
dsFuw03	1864	4.2	7	4.4	50	3906	4.6	8	3.1	36
dsFuw04	2747	4.0	7	3.5	50	5272	4.3	7	2.7	36
dsFuw05	3495	3.7	7	3.1	50	6739	4.0	7	2.3	36
dsFuw06	4033	3.6	6	2.8	50	6882	3.8	7	2.2	36
dsFuw07	4341	3.4	6	2.6	50	7479	3.7	6	2.0	36
dsFuw08	4342	3.3	6	2.4	50	7680	3.5	6	1.9	36
dsFuw09	4347	3.2	6	2.4	50	7391	3.4	6	1.9	36
dsFuw10	4438	3.1	6	2.3	50	7576	3.3	6	1.8	36
dsK	43	3.1	5	14.8	50	65	4.1	6	8.3	36
dsKuw02	652	4.0	7	6.0	53	1359	4.7	9	4.3	38
dsKuw03	1785	4.2	8	4.6	50	3629	4.6	8	3.3	36
dsKuw04	2660	3.9	7	3.7	50	5063	4.3	7	2.9	36
dsKuw05	3413	3.7	6	3.2	50	6238	4.1	7	2.4	36
dsKuw06	3959	3.5	6	2.9	50	6257	3.9	9	2.4	36
dsKuw07	4297	3.4	6	2.6	50	6706	3.8	6	2.1	36
dsKuw08	4281	3.3	6	2.5	50	7049	3.6	6	2.0	36
dsKuw09	4289	3.2	6	2.5	50	6684	3.5	6	2.0	36
dsKuw10	4404	3.1	6	2.3	50	6901	3.4	6	1.9	36

For the F-writers and the dsF and dsK rule sets, the number of available condition attributes was 6, while for the M-writers there were 6 features for dsK and 5 for dsF. Taking into account both the number of variables and the number of induced rules, the meaning of the average rule length (with the smallest value preferable because it supports interpretability) is different between 3.1 for dsF and dsK and F-writers and dsFuw10 and dsKuw10. For M-writers, dsKuw10 brings the smaller average of 3.4 than dsK. For both the female and male writer datasets, the highest values of maximal rule support were detected for dsFuw02 and dsKuw02. The highest values of average support were observed for dsF and dsK, but only because significantly lower numbers of rules were inferred.

Inferred rule sets were employed as classification systems. Their performance was measured by classification accuracy. The choice of this particular evaluation measure was dictated by the conditions in which the inducers operated [20]: the classification task was binary, the classes balanced and of the same importance, with the same misclassification cost. Weighted voting was used as the strategy for resolving conflicts. Each set, induced from a particular data variant, was tested against all variants of two test sets, transformed independently (Tind) and based on discretisation models constructed for training sets (ToL). For each discrete variant, the average accuracy was calculated, and the results obtained are shown in Fig. 1. The preference for evaluation by test sets as opposed to cross-validation comes from the characteristics of the stylometric domain [2]. The latter case results in over-optimistic results and is less reliable.

Each particular variant of rule classifier was evaluated with the help of all variants of test sets, even those with different numbers of intervals constructed by unsupervised processing. This attitude allowed for investigation of which approach is most beneficial: keeping a more detailed discrete representation for

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

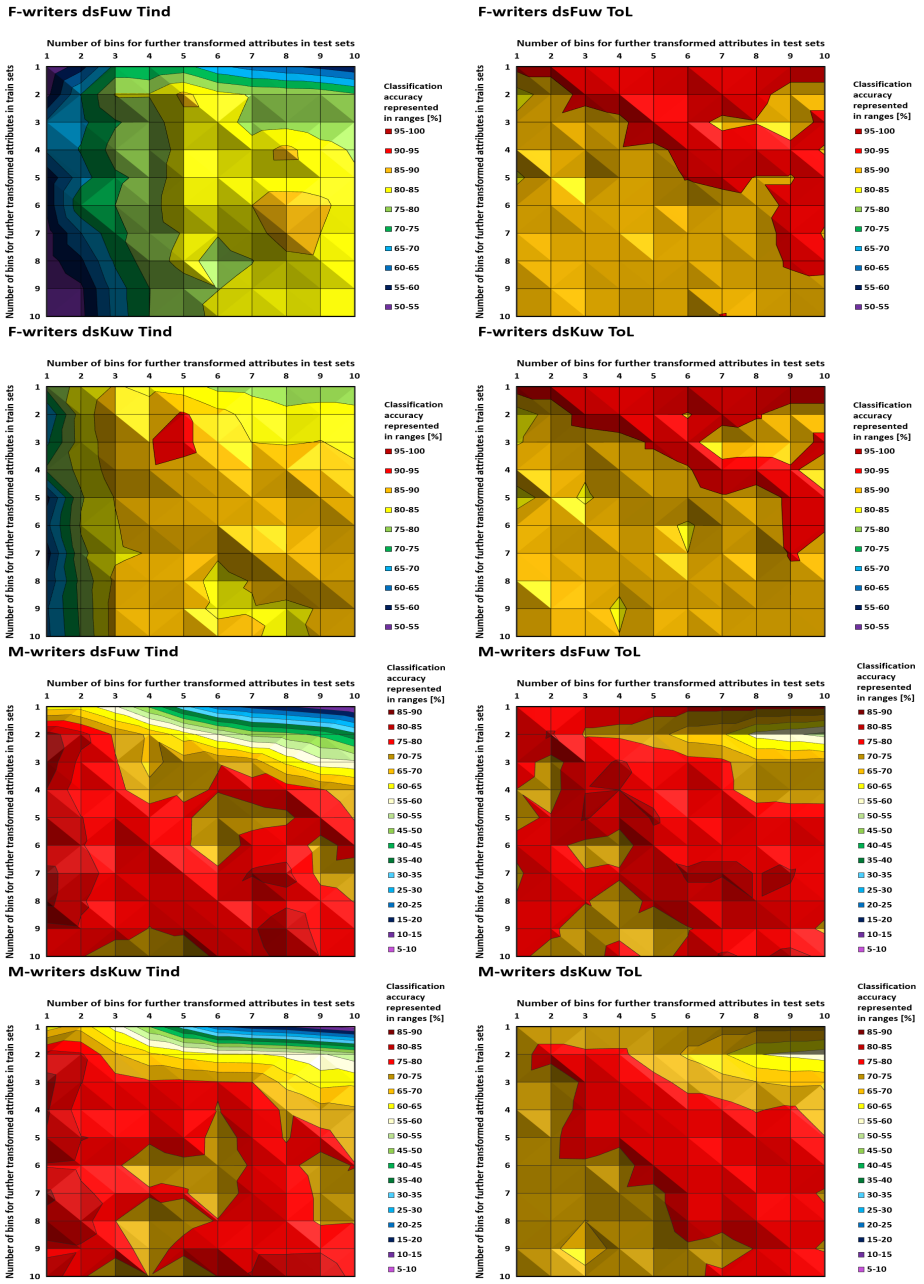

Fig. 1. Performance [%] of rule classifiers discretised by supervised Fayyad and Irani (dsF) and Kononenko (dsK) approaches in the perspective of attributes additionally transformed by unsupervised equal width binning (uw) in training and test sets, with varying bin numbers and test sets discretised independently (Tind) or based on models obtained for learning sets (ToL).

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

variables in training sets from which rules are induced, providing more elaborate definitions for test sets, or both. In addition, employing test sets obtained by various discretisation approaches widens the scope for observations of irregularities present in the data.

In the charts included in Fig. 1, the categories on the Y axis provide information on the number of intervals used in unsupervised discretisation of variables in the training sets, while the categories for the X axis do the same with respect to the test sets. The charts on the left correspond to evaluation with the test sets discretised independently and those on the right with their transformations based on definitions for intervals obtained for the respective training sets. The surface shows by colours the ranges of classification accuracy.

In the case of independently processed test sets, for the female and male writer datasets, the fusion of supervised and unsupervised discretisation procedures worked to advantage, resulting in increased performance of rule classifiers. For F-writers, the best results can be observed in the region close to the main diagonal, which corresponds to conditions of the same or close numbers of intervals used in the unsupervised discretisation procedure for both training and test sets. For discretisation by the Fayyad and Irani algorithm, the maximum can be detected for higher bin numbers than for the Kononenko algorithm. For M-writers, a different trend emerged: the best results can be detected for cases where training sets were transformed by supervised-unsupervised discretisation, while test sets were processed only by supervised algorithms. The opposite situation, with further transformed test sets used for rule classifiers obtained from dsK or dsF data variants, resulted in a very severe performance degradation.

When ToL test sets were used, in general there were more cases of advantageous accuracy than for independent discretisation. For F-writers, higher levels of correct predictions were reported when there were higher numbers of bins in test sets than in training sets subjected to unsupervised transformations. For M-writers, the highest ranges of classification accuracy are grouped more along the main diagonal (in particular, for dsK domain), which means processing with equal or close numbers of bins formed in both types of sets, training and test.

Rule-based classifiers can also be studied in the perspective of coverage provided for samples. 100% coverage means that for each instance there was some matching rule. This aspect in relation of test sets is addressed in Fig. 2. For all combinations of transformations, the coverage is shown by values and through the colour scale. The green reflects the most preferable cases, that is, the coverage 100%. When the coverage was lower, the values were indicated by white changing into shades of red. For the most part, perfect coverage was provided. The conditions where it was imperfect happened rather for independently processed test sets, and when the training sets were subjected just to supervised discretisation, or with low numbers of bins in unsupervised transformations, while for the test sets higher numbers of intervals were defined in extended processing.

The extensive experiments performed show the merits of adapting the fusion of supervised and unsupervised discretisation procedures to data in the case of supervised processing returning reduced attribute sets. Two-step transformation

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

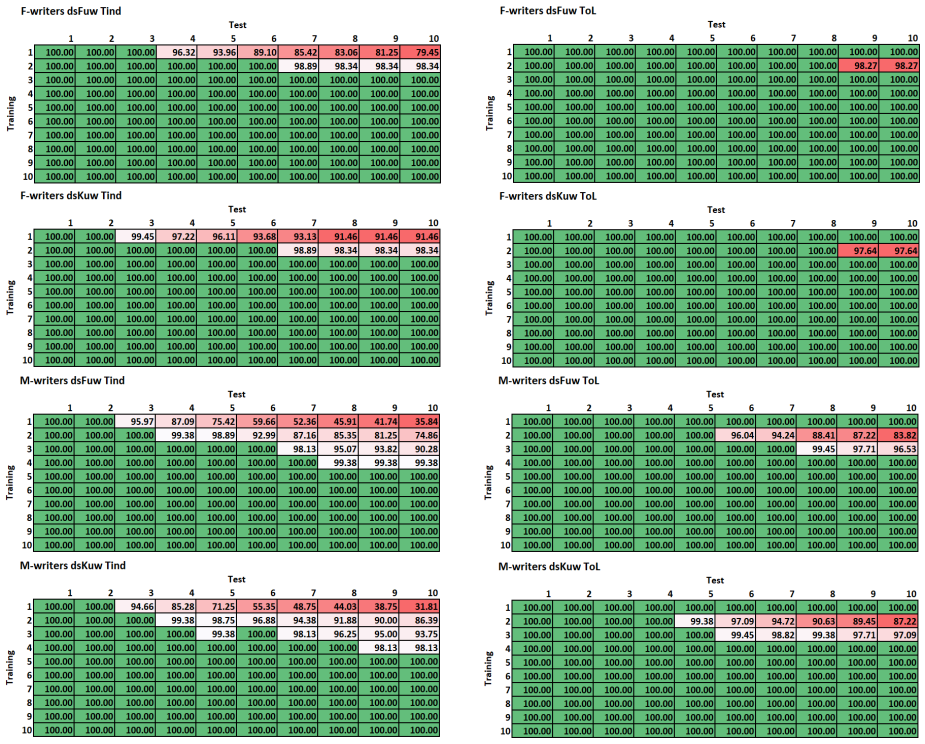

Fig. 2. Coverage [%] of rule classifiers provided for test sets discretised by supervised Fayyad and Irani (dsF) and Kononenko (dsK) approaches in the perspective of attributes additionally transformed by unsupervised equal width binning (uw) in training and test sets, with varying bin numbers and test sets discretised independently (Tind) or based on model obtained for learning sets (ToL).

ensures that all input features retain some informative content and are represented meaningfully in a discrete domain. This affects the knowledge patterns discovered and leads to improved predictions from rule-based classifiers.

4 Conclusions

In the paper, the issues related to preserving information in the process of discretisation are considered. Attributes, which may be treated as not contributing any information as a result of supervised transformations, could be completely removed from considerations. Instead, the translation process for such features is extended by unsupervised algorithms, causing fusion of supervised and unsupervised discretisation methods. The effects of this two-step procedure were evaluated with the help of approaches based on rough set theory, including reducts and decision rules. These methods discover knowledge patterns from the data and its representation. Their characteristics allow for evaluation of the transparency

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

and interpretability of the results. Experiments on two-step discretisation were conducted in two modes, when the training and test sets were discretised independently and when interval definitions for the training sets were used to find a discrete representation for the test sets.

The results from the experiments carried out in the stylometric domain show the advantages of the presented approach, visible in the increased performance of classifiers. In consequence of the fusion of supervised with unsupervised construction of intervals, all condition attributes kept some informative content, which turned out to be beneficial to characteristics and properties of knowledge patterns represented in a discrete domain, discovered by rough set processing.

The investigations described in the paper offer many directions for future research. One of these most obvious ones is the study of other discretisation algorithms from the unsupervised category. Also, application of two-step discretisation in ranking construction can be considered. Another path could involve changing the order of processing steps: performing first exploration of input continuous data and then subjecting discovered patterns to discretisation.

Acknowledgments. The research works presented in the article were carried out within the statutory project of the Department of Computer Graphics, Vision and Digital Systems (RAU-6, 2025), at the Silesian University of Technology, Gliwice, Poland, and at the Institute of Computer Science, University of Silesia in Katowice, Sosnowiec, Poland.

References

1. Alateeq, M., Pedrycz, W.: Logic-oriented fuzzy neural networks: A survey. Expert Systems with Applications **257**, 125120 (2024)
2. Baron, G.: Comparison of cross-validation and test sets approaches to evaluation of classifiers in authorship attribution domain. In: Czachórski, T., Gelenbe, E., Grochla, K., Lent, R. (eds.) Computer and Information Sciences: 31st International Symposium, ISCIS 2016, Kraków, Poland, October 27–28, 2016, Proceedings, pp. 81–89. Springer International Publishing, Cham (2016)
3. Baron, G., Harężlak, K.: On approaches to discretization of datasets used for evaluation of decision systems. In: Czarnowski, I., Caballero, A.M., Howlett, R.J., Jain, L.C. (eds.) Intelligent Decision Technologies 2016, pp. 149–159. Springer International Publishing, Cham (2016)
4. Bazan, J., Szczuka, M.: The rough set exploration system. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets III, Lecture Notes in Computer Science, vol. 3400, pp. 37–56. Springer, Berlin, Heidelberg (2005)
5. Bergamin, L., Polato, M., Aiolli, F.: Improving rule-based classifiers by Bayes point aggregation. Neurocomputing **613**, 128699 (2025)
6. Fayyad, U., Irani, K.: Multi-interval discretization of continuous valued attributes for classification learning. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence. vol. 2, pp. 1022–1027. Morgan Kaufmann (1993)
7. Garcia, S., Luengo, J., Saez, J., Lopez, V., Herrera, F.: A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. IEEE Transactions on Knowledge and Data Engineering **25**(4), 734–750 (2013)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15

8. González-Díaz, Y., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Lazo-Cortés, M.S.: Algorithm for computing all the shortest reducts based on a new pruning strategy. Information Sciences **585**, 113–126 (2022)
9. Grzegorowski, M., Slezak, D.: On resilient feature selection: Computational foundations of r-c-reducts. Information Sciences **499**, 25–44 (2019)
10. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Morgan Kaufmann, Boston (2011)
11. Jensen, R., Shen, Q.: Finding rough set reducts with ant colony optimization. In: Proceedings of the 2003 UK Workshop on Computational Intelligence. pp. 15–22 (2003)
12. Kononenko, I.: On biases in estimating multi-valued attributes. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence IJCAI'95. vol. 2, pp. 1034–1040. Morgan Kaufmann Publishers Inc., Montreal, Canada (1995)
13. Koppel, M., Schler, J., Argamon, S.: Authorship attribution: what's easy and what's hard? Journal of Law and Policy **21**(2), 317–331 (2013)
14. Pawlak, Z.: Rough sets and intelligent data analysis. Information Sciences **147**, 1–12 (2002)
15. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Information Sciences **177**(1), 3–27 (2007)
16. Rissanen, J.: Modeling by shortest data description. Automatica **14**(5), 465–471 (1978)
17. Stańczyk, U.: Application of rough set-based characterisation of attributes in feature selection and reduction. In: Virvou, M., Tsihrintzis, G.A., Jain, L.C. (eds.) Advances in Selected Artificial Intelligence Areas, Learning and Analytics in Intelligent Systems, vol. 24, chap. 3, pp. 35–55. Springer (2022)
18. Stańczyk, U.: Pruning decision rules by reduct-based weighting and ranking of features. Entropy **24**(1602), 1–28 (2022)
19. Stańczyk, U., Zielosko, B.: Assessing quality of decision reducts. In: Cristani, M., Toro, C., Zanni-Merk, C., Howlett, R.J., Jain, L.C. (eds.) Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 24rd International Conference KES-2020, Verona, Italy, 16-18 September 2020, Procedia Computer Science, vol. 176, pp. 3273–3282. Elsevier (2020)
20. Stąpor, K., Ksieniewicz, P., García, S., Woźniak, M.: How to design the fair experimental classifier evaluation. Applied Soft Computing **104**, 107219 (2021)
21. Wróbel, L., Sikora, M., Michalak, M.: Rule quality measures settings in classification, regression and survival rule induction — an empirical approach. Fundamenta Informaticae **149**, 419–449 (2016)
22. Wu, H., Zhang, Z., Wu, Q.: Exploring syntactic and semantic features for authorship attribution. Applied Soft Computing **111**, 107815 (2021)
23. Zielosko, B., Piliszczuk, M.: Greedy algorithm for attribute reduction. Fundamenta Informaticae **85**(1-4), 549–561 (2008)
24. Zielosko, B., Stańczyk, U.: Reduct-based ranking of attributes. In: Cristani, M., Toro, C., Zanni-Merk, C., Howlett, R.J., Jain, L.C. (eds.) Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 24rd International Conference KES-2020, Verona, Italy, 16-18 September 2020, Procedia Computer Science, vol. 176, pp. 2576–2585. Elsevier (2020)
25. Zielosko, B., Tetteh, E.T., Hunchak, D.: Multi-heuristic induction of decision rules. In: Campagner, A., Lenz, O.U., Xia, S., Slezak, D., Was, J., Yao, J. (eds.) Rough Sets - International Joint Conference, IJCRS 2023, Krakow, Poland, October 5-8, 2023, Proceedings. Lecture Notes in Computer Science, vol. 14481, pp. 18–30. Springer (2023)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_15