On Effectiveness of Rule Classifiers under Transformations of Input Domain

Urszula Stańczyk[0000−0002−5071−7187] and Grzegorz Baron[0000−0001−8613−631X]

Department of Computer Graphics, Vision and Digital Systems,
Silesian University of Technology, Akademicka 2A, 44-100 Gliwice, Poland
{urszula.stanczyk,grzegorz.baron}@polsl.pl

Abstract. Decision support systems often rely on the induction of decision rules, which results from the explicit presentation of premises that provide a basis on which decisions are made, improving the interpretability of the process. Algorithms for inferring rules can vary in effectiveness and in the forms in which the rule conditions are expressed. Any modifications of representation affect knowledge patterns present, therefore, they reflect also on data mining processes. The paper presents research on performance of rule classifiers constructed by the MODLEM algorithm that is capable of inducing rules from continuous data due to inherent discretisation-like processing of attribute domains. This mechanism was combined with discretisation of the input space, and the results were investigated. The experiments carried out on the task of authorship attribution led to the conclusion that a partial transformation of the input datasets can return representations of attributes that improve accuracy.

Keywords: Rule-based classifier · Rule induction · MODLEM · Discretisation · Ranking · Stylometry.

1 Introduction

High effectiveness belongs to primary objectives when decision support systems are designed and various data sources and forms are explored in search of representations advantageous to knowledge discovery approaches. When the process is characterised by transparency, it supports understanding and generalisation of learnt patterns, adaptation to changing conditions. These properties can play an important role, in particular, when many criteria are taken into account [11].

When learning from examples is performed, a closer study of the input space and characteristics of features [6], can result in improved accuracy, which is a frequent aim of investigations. Transformations such as discretisation [1] enable a wider scope of data mining techniques because not all algorithms can operate on continuous-valued attributes. Even when they can, translation into discrete type leads to simplification, which can prove to be favourable for performance.

MODLEM is a decision rule induction algorithm [10] that can work directly in continuous space. This is possible due to processing of attribute domains that closely follows discretisation principles by locating threshold values partitioning

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_24

the space [3], used to formulate rule descriptors. In the research presented in the paper, the effectiveness of MODLEM was studied under changing representation of space, caused by sequential discretisation by selected methods.

The experiments were conducted on the datasets prepared for the stylometric task of authorship attribution [5]. The nature of stylometric space explored with style markers resulted in continuous values of attributes based on text samples labelled by their authors. With uncertainty of knowledge, the data were suitable for the application of rough set approaches and the MODLEM algorithm. Observations of the results led to the conclusion that, despite the operation mode of MODLEM, direct discretisation of space can be beneficial for performance, which confirmed the merits of the methodology illustrated in the paper.

The content of the paper is organised as follows. Section 2 presents the research background. Section 3 details of the experimental setup. Section 4 is dedicated to the results obtained. Conclusions are given in Section 5.

2 Background and Related Works

The reported research involved feature selection mechanisms driving the transformation of the input domain, and their influence on the performance of constructed rule classifiers. The section presents the elements of this background.

2.1 Algorithms for Induction of Decision Rules

The MODLEM algorithm infers a minimal set of rules for every class by sequential covering. Its specific mode of operation lies in direct handling of attribute domains, in the processing similar to discretisation, without the transformation taking place. Rule induction starts with sorting values of attributes in non-decreasing order. For each condition attribute, its values are analysed to find a cut-point located in the middle between two successive values characterising examples with different class labels. The quality of this threshold is evaluated by the chosen measure [3], which can be the class entropy or the Laplacian accuracy Acc_L:

$$Acc_L = (N_C + 1)/N_{tot} + k. \tag{1}$$

Here, N_C stands for the number of positive examples covered by the condition, N_{tot} is the total number of examples covered, and k is the number of decision classes. In evaluation, higher accuracy values are preferred. Once the optimal value v_a is chosen, it is used as a condition in the premise of the rule.

The premise of a decision rule includes a conjunction of P elementary conditions (rule descriptors) listing the threshold values v_{a_i} of attributes a_i:

$$If(a_{i_1}(x)\ rel_1\ v_{a_{i_1}}) \wedge \ldots \wedge (a_{i_P}(x)\ rel_P\ v_{a_{i_P}})\ then\ v_d, \tag{2}$$

and v_d is the label of the class to which an example should be assigned when all conditions are met. The forms used for the conditions depend on a relational operator rel. The selected v_a is incorporated in a condition on a as either $(a < v_a)$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_24

or $(a \geq v_a)$, depending on which relation covers more examples. For a rule, some attribute can be chosen with two threshold points v_1 and v_2, leading to a condition $a = [v_1, v_2)$. This descriptor is a consequence of forming an intersection of conditions $(a < v_2)$ or $(a \geq v_1)$ for $v_1 < v_2$. The simplest descriptors are defined for nominal or discrete attributes because then $(a = v_a)$ can be used.

2.2 Characteristics of the Input Space and Transformations

Data are at the centre of any exploration procedures applied. The type and amount available determine how it can be processed. Understanding domain characteristics is one of the key elements in the construction of an effective classification system. Recognition of the importance of features is another factor of such an impact. Relevance assessment begins with expert domain knowledge, but feature selection techniques can support the process [6]. One of such mechanisms is ranking, which orders the attributes from most to least relevant.

When the input domain is transformed, the patterns and characteristics of the attributes also undergo some changes. Discretisation procedure is responsible for replacing continuous values of features with categorical representations [1]. The domains of the variables are partitioned into intervals, and for each a nominal value is defined. Two main groups of algorithms are distinguished: supervised and unsupervised [7]. In supervised discretisation, information on classes is relevant to the interval construction problem. Unsupervised methods focus exclusively on the transformed domain and observed attribute values.

Discretisation is commonly included in the pre-processing stage. So, the data are first transformed and then explored. If the learnt knowledge patterns can be directly accessible (as in the case of decision rules), the order of proceeding can be reversed, and the knowledge discovered first and then discretised [9]. Typically, one algorithm is applied to all domains, transformed at the same time.

2.3 Motivation

The MODLEM algorithm does not need discrete data to operate. However, it does not imply that it is impervious to transformation of the input domain. The observed degree of sensitivity to data processing depends on the type of processing and the specific characteristics of the input space under investigation. This line of reasoning was the primary motivation for the presented research.

Furthermore, no single discretisation algorithm is optimal for all conditions. Supervised approaches are often considered superior to unsupervised ones due to their support for class distinction. For both categories of methods, several variants were developed, making the choice not trivial and suggesting a study in search of advantageous representation, focused on properties of a specific domain.

3 Experimental Setup

In the investigations, WEKA software [4] was applied for data transformations and knowledge discovery. The section details the scope of the experiments.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_24

3.1 Application Domain and Input Features

The task chosen for the experiments was binary authorship attribution, from the domain of stylometric analysis of texts, based on writing styles [5]. The authors studied were two pairs of known writers, Edith Wharton and Mary Johnston, and Henry James and Thomas Hardy. Their literary works, divided into smaller text blocks of comparable size, provided the text corpus for analysis.

The selected attributes reflected the frequency of occurrence of 12 function words [12]. This made the features continuous. They were calculated over text samples and grouped into three sets included in each of the two datasets (the female writer dataset, F-writers, and the male writer dataset, M-writers): one training set and two test sets. In all sets, the data was balanced.

Three orderings of attributes were obtained, shown in Table 1. Two were returned from the ranking mechanism: WrapB which belongs to the wrapper category and Relief, which is instance-based [6]. In addition, over the samples included in the train sets, the average values of all features were calculated, and then the attributes were sorted by this value, resulting in the third ordering. The three orders were used to control sequential discretisation of space.

Table 1. Relief and WrapB ranking and average-based ordering of attributes

Position	F-writers												Order	M-writers											Position
	1	2	3	4	5	6	7	8	9	10	11	12		1	2	3	4	5	6	7	8	9	10	11	12	
	on	to	of	as	by	if	or	up	at	in	so	no	Relief	by	if	so	or	in	as	at	on	no	of	up	to	
	to	on	of	no	at	if	so	up	in	or	by	as	WrapB	by	if	to	in	so	no	at	of	as	on	up	or	
	of	to	in	as	at	on	by	so	no	if	or	up	Average	to	of	in	as	at	on	so	by	if	no	up	or	

3.2 Discretisation Approaches and Methodology of Transformations

Four main discretisation methods were applied to the data. Unsupervised equal width (duw) binning defines the required number of bins with equal width. Unsupervised equal frequency (duf) binning constructs such intervals that represent the same number of original datapoints. For both methods, the number of bins ranged from two to ten. Fayyad and Irani (dsF) [2] and Kononenko (dsK) [7] are supervised algorithms that refer to the Minimum Description Length principle. All in all, the total number of discretisation approaches was 20, but more data versions were obtained due to the methodology of transformations implemented.

The input data was processed gradually, one attribute at a time. The feature to be translated was selected on the basis of an ordering provided as the input parameter to the procedure, starting with the highest positions in the ordering and then sequentially proceeding down the list of variables. With three orderings used and 12 features, the total number of data variants studied was equal to

$1 \text{ continuous} + (11 \times 3) \times 20 \text{ partially discrete} + 20 \text{ entirely discrete} = 681$

All data variants were explored and decision rules inferred from the training sets by the MODLEM algorithm with Laplacian accuracy used as a measure when conditions were evaluated. The performance was then verified by classifying

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_24

samples from the corresponding test sets. The test sets were discretised using definitions of intervals formed in transformations of the training sets [8].

4 Results From Experiments

The investigations began with establishing inducer effectiveness in the continuous domain, expressed by the classification accuracy averaged over the test sets. For the female writer dataset, the accuracy was 92.08% and for the male writer dataset it was 79.31%. These two values were treated as reference points.

The results of the subsequent rule induction under changing conditions are shown in Table 2 by differences, with the reference classification accuracy subtracted from the performance reported in each case. Positive values correspond to improvement, and negative values indicate the degraded power of the classifiers.

For the F-writer dataset, when a few bins were constructed by unsupervised methods, a noticeable drop in performance was noted. With increasing numbers of discretised features, the predictions were closer to the reference point. duw presented fewer cases of improved accuracy than duf. Supervised discretisation methods returned observations mostly to the disadvantage. The highest number of favourable cases occurred for the average-based order of variables. The rankings produced close results, with a slight superiority of WrapB over Relief.

For M-writers, when two bins were defined by two unsupervised approaches and WrapB or Relief directed the processing, only degraded powers of classifiers were noted. In other processing paths, for some number of discretised features, such classification accuracy was recorded which was an improvement over the reference point in the continuous domain. Supervised discretisation resulted in more cases of enhanced predictions than for F-writers. With the exception of a pairing Relief with duf, equal width binning gave better results than following distributions when WrapB or average-based orders were involved in processing.

For all transformations paths, the statistics were calculated, referring to the entire sequential discretisation path (from a single categorised variable to the entirely discrete space). They are provided by Table 3 and include the average classification accuracy with the standard deviation and the minimum and maximum accuracy. Among all discrete variants, in each row the preferred best values (highest for classification accuracy, lowest for *std*) are marked in bold font.

In general, the values of *std* were higher mainly when a low number of bins was constructed for variables by unsupervised approaches. The lowest values were reported for more bins defined for attributes. For the three orderings controlling transformations, the statistics for two supervised algorithms were relatively close and inferior to those for unsupervised discretisation.

For F-writers, the best scenario for rule induction was for equal frequency binning with six bins and following WrapB ranking. For M-writers, the same ordering and discretisation approach worked the best, but for ten bins. For both datasets, the highest minimum was found in the average-based order, but it was not accompanied by the advantageous maximum or average accuracy. However, for M-writers the best maximum was found for this order and duf4 approach.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_24

Table 2. Difference in performance of rule classifiers observed in sequential discretisation following *Order*, with P giving the number of transformed features. The discretisation approaches included supervised methods (dsF for Fayyad and Irani, and dsK for Kononenko), and unsupervised equal frequency (duf) and equal width (duw) binning.

Order	P	Supervised		Unsupervised duf									Unsupervised duw								
		dsF	dsK	2	3	4	5	6	7	8	9	10	2	3	4	5	6	7	8	9	10
P-writers																					
WrapB	1	-1.667	-1.667	-2.222	1.875	0.833	0.208	-0.347	-1.111	0.833	0.833	0.278	0.764	-2.847	-0.972	-0.833	1.389	-2.153	-0.417	-0.972	0.278
	2	0.139	0.139	-12.153	2.014	-0.278	-2.361	1.389	-0.486	-0.347	-0.486	-0.833	-5.833	-3.403	-1.042	-2.778	-0.556	0.278	-2.153	0.208	-2.847
	3	-6.250	-3.264	-8.472	-0.972	-0.347	1.181	1.944	0.694	-1.042	2.569	-2.014	-5.764	-3.125	0.069	-6.250	-1.597	1.458	1.389	2.014	-1.042
	4	-8.681	-3.889	-8.472	-0.417	1.389	1.875	1.319	0.694	0.000	2.500	-1.458	-5.764	-2.639	0.069	-6.250	-1.597	1.528	0.833	2.014	-1.042
	5	-1.528	-1.667	-4.583	-2.292	-0.278	1.875	3.125	-1.597	0.069	1.250	-1.458	-5.139	-5.486	-1.111	-2.708	-3.403	-0.278	0.833	2.014	-0.417
	6	-1.528	-1.667	-4.028	-1.736	-0.833	2.500	3.125	-1.597	-0.486	1.875	-1.458	-5.833	-5.486	-0.486	-3.333	-2.778	0.278	0.833	2.014	-0.417
	7	-0.347	0.139	-4.514	-1.667	1.875	0.694	3.750	-1.597	-1.042	1.875	-2.083	-5.833	-4.931	-1.042	-4.444	-2.778	-1.528	-0.903	1.458	-0.417
	8	-0.972	-0.417	-2.153	-1.667	1.875	0.069	1.944	-1.042	1.250	1.875	-3.264	-5.833	-8.681	-1.042	0.278	-2.778	-1.528	0.208	1.458	0.139
	9	-0.347	-1.042	-3.333	-1.597	1.250	2.986	1.944	0.694	1.250	2.500	-3.264	-5.208	-3.333	-1.736	0.278	-2.153	-2.083	-0.972	0.833	-0.972
	10	-2.708	-4.583	-0.903	-2.847	0.139	1.875	1.319	-0.486	1.250	1.944	-3.264	-6.944	-3.958	-0.486	-0.903	-2.153	-2.708	-0.972	-0.347	-0.972
	11	-6.250	-3.889	-3.333	-2.292	-1.111	2.500	0.139	0.139	1.250	1.875	-5.139	-4.722	-5.000	0.139	-2.917	-1.597	0.278	-0.972	0.833	-1.042
	12	-1.181	-2.361	-5.694	-0.556	2.500	0.069	1.806	3.681	-0.486	-0.417	2.014	-7.014	-2.847	-4.028	1.389	-1.667	-0.486	-1.042	-1.528	-1.597
Relief	1	-1.111	-1.111	-3.333	0.764	-1.528	0.694	-2.014	-1.042	-0.486	-1.597	-0.903	-1.667	-4.097	-1.042	-1.042	-3.958	-0.417	-2.639	-1.597	-2.153
	2	0.139	0.139	-12.153	2.014	-0.278	-2.361	1.389	-0.486	-0.347	-0.486	-0.833	-5.833	-3.403	-1.042	-2.778	-0.556	0.278	-2.153	0.208	-2.847
	3	-6.250	-3.264	-8.472	-0.972	-0.347	1.181	1.944	0.694	-1.042	2.569	-2.014	-5.764	-3.125	0.069	-6.250	-1.597	1.458	1.389	2.014	-1.042
	4	-8.194	-7.500	-8.542	-0.347	2.014	0.694	0.833	1.875	1.389	-2.014	0.278	-12.361	-4.931	1.458	-2.153	-3.264	0.208	-0.972	-0.417	-0.556
	5	-6.944	-4.028	-4.444	0.139	1.458	2.500	0.833	1.875	1.389	-2.639	-0.278	-7.083	-1.528	-2.292	-1.597	0.139	-0.347	-0.417	-0.417	-1.181
	6	-7.569	-4.653	-4.097	0.764	0.903	3.681	1.389	1.875	0.833	0.903	-0.278	-8.750	-1.528	-1.250	-2.847	0.139	-0.347	-0.417	-0.417	0.556
	7	-6.528	-3.611	-6.875	-1.042	-0.903	3.125	1.250	1.875	0.833	0.903	1.389	-8.681	-0.903	-1.181	-1.042	-0.417	-0.417	-0.417	-0.417	0.556
	8	-7.153	-4.167	-5.278	-0.417	-0.903	2.431	1.250	0.833	1.389	0.903	1.389	-4.653	-1.528	-1.181	-0.417	-2.778	0.139	1.319	-0.417	0.556
	9	-6.458	-2.917	-5.208	-1.181	-0.417	1.250	1.806	-4.028	0.208	-1.458	1.389	-5.625	-5.000	-1.806	-0.417	-3.333	-1.597	1.319	-0.972	0.000
	10	-1.111	0.625	-2.292	-1.181	0.139	1.250	0.694	1.181	-0.417	1.319	1.389	-4.444	-0.417	-2.431	0.764	-0.417	-1.597	-0.417	-0.972	-1.597
	11	-2.986	-3.542	-5.833	0.069	1.389	1.875	0.694	2.431	1.944	0.764	1.389	-4.583	-0.486	-4.028	0.764	-1.042	-0.486	-1.042	-1.528	-1.042
	12	-1.181	-2.361	-5.694	-0.556	2.500	0.069	1.806	3.681	-0.486	-0.417	2.014	-7.014	-2.847	-4.028	1.389	-1.667	-0.486	-1.042	-1.528	-1.597
Average	1	3.056	3.125	2.014	-0.417	1.389	1.319	3.056	1.250	0.764	0.139	-0.972	2.569	2.431	1.458	0.903	1.458	2.500	3.750	3.194	-0.278
	2	-3.889	-0.903	-5.139	-0.486	0.278	1.319	3.056	0.764	1.389	0.139	0.208	0.139	-0.903	1.250	-2.014	2.569	2.014	3.056	2.639	-2.639
	3	-2.153	-1.528	-4.653	-2.153	0.278	0.278	2.500	1.319	-1.458	-2.153	-2.708	-1.667	0.833	0.625	-2.708	2.569	2.014	2.431	2.639	-2.083
	4	-6.319	-3.958	-1.389	-3.333	0.903	0.903	2.500	-0.903	-0.347	1.319	-2.708	-3.403	-1.528	-1.111	-1.528	-0.972	1.389	1.250	1.389	-2.153
	5	-5.833	-3.472	0.833	-3.958	0.903	0.278	1.319	0.139	0.208	0.139	-2.708	-1.597	-1.458	-0.486	-0.278	-1.597	0.833	-0.486	1.389	-2.153
	6	-4.653	-4.028	-5.625	-1.181	1.389	-1.528	0.833	-0.417	0.208	-2.708	0.278	-9.931	-1.597	-0.972	0.139	-1.597	-1.597	-1.528	-0.347	-3.403
	7	-1.181	-0.069	-4.583	-0.556	0.764	1.875	0.208	0.694	0.764	-2.153	-0.278	-10.347	-0.278	0.139	0.139	-1.111	-2.153	-2.153	-0.347	-2.222
	8	-0.556	0.139	-5.625	-1.111	0.208	1.319	0.208	0.694	1.389	-2.778	-0.278	-9.722	-2.222	-2.917	2.569	-2.292	-1.042	-2.153	-0.347	-2.222
	9	-2.222	-1.597	0.278	0.069	1.458	1.875	0.625	1.250	0.833	-1.111	0.903	-7.431	-6.250	-4.653	2.569	-2.292	-1.042	-2.153	-0.347	-4.028
	10	-2.292	-1.736	-3.194	0.069	0.764	1.944	1.806	1.250	1.389	0.764	0.903	-8.542	-1.597	-4.722	0.764	-2.292	-1.597	-2.153	-0.347	-3.472
	11	-4.028	-2.847	-2.778	-3.542	1.944	1.319	1.806	3.056	-0.486	0.764	1.389	-5.625	-1.042	-5.278	1.389	-2.292	-1.042	-2.778	-1.528	-3.472
	12	-1.181	-2.361	-5.694	-0.556	2.500	0.069	1.806	3.681	-0.486	-0.417	2.014	-7.014	-2.847	-4.028	1.389	-1.667	-0.486	-1.042	-1.528	-1.597
M-writers																					
WrapB	1	-2.222	-2.222	-0.069	-3.542	-2.500	-1.944	-2.153	-4.653	-5.556	-0.556	-0.417	-5.764	-1.042	-5.208	-3.056	-2.083	-2.292	-2.847	-2.431	-2.431
	2	-0.972	-0.972	-4.444	-1.736	-0.139	-0.278	1.319	-2.778	-4.306	0.069	0.625	-4.722	-0.903	-4.375	-3.472	0.694	-1.667	-3.403	-1.806	-1.181
	3	-3.472	-3.472	-3.194	-8.056	-2.292	-1.319	0.694	-2.222	-3.542	-0.556	-0.417	-5.833	1.458	-3.125	-2.361	-1.042	-3.958	-2.083	-3.403	-1.181
	4	-2.708	-2.708	-4.514	-2.708	1.111	-1.250	3.125	-1.667	-1.875	-0.972	1.875	-5.000	-0.278	-1.597	-1.111	1.250	0.208	-0.972	-1.111	0.069
	5	1.250	1.250	-4.514	-3.333	1.736	0.625	0.208	0.000	-1.319	0.278	2.431	-0.208	-1.667	0.208	-1.111	2.986	-1.597	0.208	0.069	0.069
	6	-3.264	-3.264	-2.708	0.417	3.958	-0.139	1.250	0.069	0.903	-3.542	3.611	-0.833	1.528	1.250	-2.292	0.694	0.139	3.056	0.625	0.764
	7	-4.306	-4.306	-3.125	-1.875	3.958	2.708	3.472	1.597	0.903	3.125	3.403	-5.278	-0.069	0.069	-0.694	-1.250	1.319	3.542	-1.250	-1.250
	8	0.972	-0.833	-5.417	1.667	-0.208	0.347	-1.806	-0.208	-0.139	-1.250	3.264	-4.514	-0.069	0.000	-0.139	0.347	0.417	3.403	2.292	-3.056
	9	-5.486	-4.306	-4.167	2.361	-0.764	2.153	-2.361	0.417	0.417	-0.486	3.819	-2.153	1.111	0.417	-2.639	-1.319	1.250	1.042	1.597	1.111
	10	-3.056	-3.056	-5.972	2.917	-0.417	0.417	-0.625	0.972	1.667	1.806	2.847	-2.847	-0.139	-3.542	-2.639	2.708	2.292	2.986	-1.250	0.000
	11	-3.681	-6.528	-7.569	-3.611	2.222	0.069	-0.764	1.042	0.972	1.806	4.653	-4.653	-0.694	1.528	-1.806	-0.764	2.292	-0.833	-2.361	1.806
	12	-2.292	-4.167	-11.111	-0.486	0.347	-0.069	0.000	2.708	1.736	0.556	0.347	-5.764	-2.569	1.111	-1.181	-3.542	-0.486	0.486	0.417	1.111
Relief	1	-2.222	-2.222	-0.069	-3.542	-2.500	-1.944	-2.153	-4.653	-5.556	-0.556	-0.417	-5.764	-1.042	-5.208	-3.056	-2.083	-2.292	-2.847	-2.431	-2.431
	2	-0.972	-0.972	-4.444	-1.736	-0.139	-0.278	1.319	-2.778	-4.306	0.069	0.625	-4.722	-0.903	-4.375	-3.472	0.694	-1.667	-3.403	-1.806	-1.181
	3	1.181	1.181	-4.514	-5.694	1.667	-0.903	1.319	-0.625	-4.306	0.139	0.556	-1.806	-5.208	-4.236	-2.361	1.250	-1.667	-5.833	-0.625	-1.181
	4	-0.833	-0.833	-2.500	0.625	1.736	0.417	0.764	0.486	-2.500	0.625	2.778	-4.236	-0.625	-0.833	1.181	-1.806	-3.611	-0.764	-0.556	1.667
	5	0.000	0.000	-5.833	5.486	2.917	-0.764	4.167	-1.736	-0.069	1.319	2.778	-6.250	0.208	0.069	-0.625	-0.417	-1.250	1.875	0.625	1.667
	6	-0.694	-0.694	-4.653	2.569	2.361	2.222	2.500	-2.431	-1.250	1.250	2.708	-11.042	-0.903	-4.306	0.000	-1.111	-0.486	0.625	1.736	2.917
	7	-0.139	-0.139	-3.125	-2.917	1.528	3.194	1.528	-2.569	-0.139	-0.694	2.778	-5.833	-9.375	-3.681	-1.389	-1.806	0.069	-0.208	1.667	1.597
	8	-6.736	-6.736	-4.167	-5.208	-0.278	2.222	-0.208	2.222	1.181	-0.069	1.042	-8.819	-1.875	-2.500	-1.875	-2.986	0.069	-0.764	2.361	-0.069
	9	-4.861	-4.861	-4.514	-1.667	1.458	2.778	-0.069	2.014	0.625	-4.306	2.778	-5.694	2.917	-1.736	-3.056	-0.764	-0.069	-0.833	1.111	0.000
	10	-3.125	-3.819	-2.569	-0.625	1.389	1.528	1.736	0.903	-0.694	0.903	2.014	-5.278	1.042	0.069	-5.417	-0.069	-0.694	-0.833	3.403	-0.694
	11	-6.875	-6.944	-4.306	-4.653	4.444	1.806	1.806	1.528	1.667	0.486	2.639	-9.861	1.528	-2.292	-4.653	1.736	-1.319	-1.389	1.667	-0.069
	12	-2.292	-4.167	-11.111	-0.486	0.347	-0.069	0.000	2.708	1.736	0.556	0.347	-5.764	-2.569	1.111	-1.181	-3.542	-0.486	0.486	0.417	1.111
Average	1	1.250	1.250	1.250	0.139	1.875	1.250	0.000	0.625	1.250	0.000	-0.556	1.250	0.139	0.139	1.250	0.069	0.625	0.000	0.625	1.250
	2	-0.139	-0.625	-0.625	-0.694	0.625	-0.694	0.417	0.556	0.000	1.111	-0.694	-1.736	-0.694	1.042	1.597	-0.625	0.556	0.972	-0.139	1.736
	3	-0.556	-0.556	-1.111	-0.556	-1.736	0.069	0.694	-1.875	-0.625	3.333	2.222	1.181	0.625	1.042	1.667	-0.069	0.625	2.222	4.583	2.917
	4	1.042	0.417	-0.139	-0.694	0.833	1.181	-0.625	2.153	2.778	-0.139	1.042	3.958	2.292	0.347	2.292	1.667	0.556	1.528	2.361	2.847
	5	0.208	1.319	-0.903	0.139	-0.347	-0.833	-0.764	-2.569	-0.625	-1.250	-0.139	0.764	-0.833	0.972	2.083	-1.944	1.736	-0.208	0.417	0.972
	6	-0.417	0.208	0.833	-2.569	-1.597	-0.278	-3.125	-4.236	-0.625	-0.694	1.111	2.083	0.417	0.972	0.972	1.458	2.292	0.486	0.347	1.667
	7	-0.833	-0.833	4.514	0.278	2.014	-2.500	2.153	0.347	2.778	-0.139	-1.181	2.014	-0.208	1.944	-1.875	2.639	0.278	2.847	-0.208	2.847
	8	0.417	0.417	-3.819	-0.347	0.903	1.597	3.958	2.014	1.806	0.486	3.403	-1.806	0.208	-1.389	-2.083	1.042	-1.597	2.361	-2.083	0.903
	9	-1.250	-1.250	-4.306	1.181	5.069	3.958	-1.319	0.278	0.486	0.625	2.778	-5.903	0.833	-1.042	-3.264	-2.153	-0.069	4.236	-0.833	-0.139
	10	-3.056	-3.056	-5.972	2.917	-0.417	0.417	-0.625	0.972	1.667	1.806	2.847	-2.847	-0.139	-3.542	-2.639	2.708	2.292	2.986	-1.250	0.000
	11	-3.681	-6.528	-7.569	-3.611	2.222	0.069	-0.764	1.042	0.972	1.806	4.653	-4.653	-0.694	1.528	-1.806	-0.764	2.292	-0.833	-2.361	1.806
	12	-2.292	-4.167	-11.111	-0.486	0.347	-0.069	0.000	2.708	1.736	0.556	0.347	-5.764	-2.569	1.111	-1.181	-3.542	-0.486	0.486	0.417	1.111

The experimental results confirmed the merits of the illustrated methodology for controlled sequential discretisation of the input space. The high number of conditions leading to improved performance of inducers provides motivation for a deeper study of attribute representation even if the knowledge mining algorithm is capable of operating in continuous space. The MODLEM algorithm is undeniably effective, yet, it can gain from transformations of the input domain.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_24

Table 3. Statistics of performance of rule classifiers observed in sequential discretisation of attributes following *Order*. Met_D details supervised (dsF for Fayyad and Irani, dsK for Kononenko) or unsupervised discretisation approach (duf for equal frequency or duw for equal width binning), *Stat* specifies statistic measures (*min*, *max* or *avg* classification accuracy, or standard deviation *std*)

							Number of bins constructed for attributes								
	Order	Met_D	Stat		Met_D	Stat	2	3	4	5	6	7	8	9	10
F-writers	WrapB	dsF	min	83.40	duf	min	79.93	89.24	90.97	89.72	**91.74**	90.49	91.04	91.60	86.94
			max	92.22		max	91.18	94.10	94.58	95.07	**95.83**	95.76	93.33	94.65	94.10
			avg	89.47		avg	87.09	91.07	92.67	93.21	**93.87**	91.92	92.29	93.60	90.25
			std	2.85		std	3.24	1.55	1.19	1.49	1.18	1.51	**0.91**	1.05	1.85
		dsK	min	87.50	duw	min	85.07	83.40	88.06	85.83	88.68	89.38	89.93	**90.56**	89.24
			max	92.22		min	92.85	89.44	92.22	93.47	93.47	93.61	93.47	**94.10**	92.36
			avg	90.07		avg	86.82	87.77	91.11	89.71	90.28	91.50	91.81	**92.92**	91.22
			std	1.61		std	2.01	1.73	1.12	2.50	1.26	1.41	1.07	1.24	**0.83**
	Relief	dsF	min	83.89	duf	min	79.93	90.90	90.56	89.72	90.07	88.06	**91.04**	89.44	90.07
			max	92.22		max	89.79	94.10	94.58	**95.76**	94.03	**95.76**	94.03	94.65	94.10
			avg	87.47		avg	86.06	91.92	92.42	**93.45**	93.07	92.98	92.52	91.98	92.49
			std	3.09		std	2.69	**0.97**	1.29	1.59	1.04	2.00	0.98	1.58	1.26
		dsK	min	84.58	duw	min	79.72	87.08	88.06	85.83	88.13	**90.49**	89.44	**90.49**	89.24
			max	92.71		max	90.42	91.67	93.54	93.47	92.22	93.54	93.47	**94.10**	92.64
			avg	89.05		avg	85.71	89.60	90.52	90.78	90.52	**91.78**	91.63	91.55	91.22
			std	2.20		std	2.71	1.65	1.54	2.07	1.44	**0.82**	1.29	0.98	1.12
	Average	dsF	min	85.76	duf	min	86.39	88.13	**92.29**	90.56	**92.29**	91.18	90.63	89.31	89.38
			max	95.14		max	94.10	92.15	94.58	94.03	95.14	**95.76**	93.47	93.40	94.10
			avg	89.48		avg	89.12	90.65	93.15	93.00	**93.73**	93.15	92.43	91.41	91.75
			std	2.57		std	2.77	1.45	**0.70**	1.00	1.02	1.29	0.90	1.45	1.64
		dsK	min	88.06	duw	min	81.74	85.83	86.81	89.38	89.79	89.93	89.31	**90.56**	88.06
			max	95.21		max	94.65	94.51	93.54	94.65	94.65	94.58	**95.83**	95.28	91.81
			avg	90.48		avg	86.87	90.71	90.36	92.36	91.29	92.07	91.75	**92.62**	89.61
			std	2.03		std	4.33	2.08	2.47	1.69	1.88	1.66	2.33	1.65	**1.02**
M-writers	WrapB	dsF	min	73.82	duf	min	68.19	71.25	76.81	77.36	76.94	74.65	73.75	75.76	**78.89**
			max	80.56		max	79.24	82.22	83.26	82.01	82.78	82.01	81.04	82.43	**83.96**
			avg	76.87		avg	74.57	77.81	79.89	79.42	79.50	78.91	78.46	79.33	**81.48**
			std	2.00		std	2.73	3.10	2.11	**1.34**	1.90	2.07	2.47	1.72	1.74
		dsK	min	72.78	duw	min	73.47	**76.74**	74.10	75.83	75.76	75.35	75.90	75.90	76.25
			max	80.56		max	79.10	80.83	80.83	79.17	82.29	81.60	**82.85**	81.60	81.11
			avg	76.42		avg	75.34	79.03	78.20	77.43	79.20	79.13	**79.69**	78.59	78.96
			std	2.02		std	1.97	1.23	2.37	**1.03**	1.92	1.91	2.48	1.73	1.49
	Relief	dsF	min	72.43	duf	min	68.19	73.61	76.81	77.36	77.15	74.65	73.75	75.00	**78.89**
			max	80.49		max	79.24	**84.79**	83.75	82.50	83.47	82.01	81.04	80.63	82.08
			avg	77.01		avg	74.99	77.82	80.55	80.16	80.36	78.89	78.17	79.28	**81.02**
			std	2.64		std	2.60	3.29	1.76	1.66	1.58	2.39	2.49	1.49	**1.20**
		dsK	min	72.36	duw	min	68.26	69.93	74.10	73.89	75.76	75.69	73.47	**76.88**	**76.88**
			max	80.49		max	77.50	82.22	80.42	80.49	81.04	79.38	81.18	**82.71**	82.22
			avg	76.79		avg	73.05	77.91	76.98	77.15	78.40	78.19	78.15	**79.94**	79.58
			std	2.72		std	2.53	3.26	2.08	1.91	1.64	**1.09**	2.05	1.72	1.54
	Average	dsF	min	75.63	duf	min	68.19	75.69	77.57	76.81	76.18	75.07	**78.68**	78.06	78.13
			max	80.56		max	83.82	82.22	**84.38**	83.26	83.26	82.01	82.08	82.64	83.96
			avg	78.53		avg	76.89	78.95	80.12	79.65	79.31	79.47	80.27	79.93	**80.63**
			std	1.55		std	4.31	1.64	1.86	1.58	1.77	2.06	**1.25**	**1.25**	1.85
		dsK	min	72.78	duw	min	73.40	76.74	75.76	76.04	75.76	77.71	78.47	76.94	**79.17**
			max	80.63		max	83.26	81.60	81.25	81.60	82.01	81.60	83.54	**83.89**	82.22
			avg	78.19		avg	78.35	79.25	79.57	79.06	79.35	80.06	80.73	79.46	**80.80**
			std	2.35		std	3.31	1.16	1.55	2.06	1.96	1.21	1.53	1.89	**1.03**

5 Conclusions

The paper presents investigations into the effectiveness of classifiers based on decision rules induced by the MODLEM algorithm. This algorithm is specific in its operation mode, as it can work directly on discrete and continuous attributes. When conditions for rules are defined, the datapoints are evaluated in a way that reminds of discretisation, but without obtaining categorical forms for descriptors.

To study the sensitivity of the rule induction procedures to the type of attributes, the sequential discretisation methodology was applied. The processing direction was controlled by the orderings of attributes. Experiments were carried out on the datasets prepared for the stylometric task of authorship attribution.

An analysis of the results led to the observations that partial discretisation of the input space can provide such conditions for the induction process, which

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_24

prove to return sets of rules more effective in labelling of unknown samples. Contrary to widely held opinions, supervised transformations turned out to be less advantageous than unsupervised algorithms. In the vast majority of discretisation paths explored, some cases of improved predictions were detected. This validated the illustrated methodology, which can be useful in cases where an effective, and at the same time transparent, path of decision making is needed.

Future works will be dedicated to comparison of performance of MODLEM-based classifiers constructed under various conditions against the operation of inducers relying on rule sets inferred by other algorithms. In addition, a comparative study will be carried out for the methodology in which rules are induced from continuous data and then the conditions are discretised.

Acknowledgments. The research presented was performed in the statutory project of the Department of Computer Graphics, Vision and Digital Systems (RAU-6, 2026).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Dash, R., Paramguru, R.L., Dash, R.: Comparative analysis of supervised and unsupervised discretization techniques. International Journal of Advances in Science and Technology **2**(3), 29–37 (2011)
2. Fayyad, U.M., Irani, K.B.: Multi-interval discretization of continuousvalued attributes for classification learning. In: 13th International Joint Conference on Articial Intelligence. vol. 2, pp. 1022–1027. Morgan Kaufmann Publishers (1993)
3. Grzymała-Busse, J., Stefanowski, J.: Three discretization methods for rule induction. International Journal of Intelligent Systystems **16**, 29–38 (01 2001)
4. Hall, M., et al.: The WEKA data mining software: an update. SIGKDD Explorations **11**(1), 10–18 (2009)
5. He, X., et al.: Authorship attribution methods, challenges, and future research directions: A comprehensive survey. Information **15**(3) (2024)
6. Kononenko, I.: Estimating attributes: Analysis and extensions of RELIEF. In: Bergadano, F., De Raedt, L. (eds.) Machine Learning: ECML-94. LNCS, vol. 784, pp. 171–182. Springer Verlag (1994)
7. Kononenko, I.: On biases in estimating multi-valued attributes. In: 14th International Joint Conference on Articial Intelligence. pp. 1034–1040 (1995)
8. Stańczyk, U., Zielosko, B.: Data irregularities in discretisation of test sets used for evaluation of classification systems: A case study on authorship attribution. Bulletin of the Polish Academy of Sciences: Technical Sciences **69**(4), 1–12 (2021)
9. Stańczyk, U., Zielosko, B., Baron, G.: Discretisation of conditions in decision rules induced for continuous data. PLOS ONE **15**(4), 1–33 (04 2020)
10. Stefanowski, J.: On combined classifiers, rule induction and rough sets. In: Peters, J.F., et al. (eds.) Transactions on Rough Sets VI: Commemorating the Life and Work of Zdzisław Pawlak, Part I, pp. 329–350. Springer, Berlin, Heidelberg (2007)
11. Więckowski, J., Sałabun, W.: Sensitivity analysis approaches in multi-criteria decision analysis: A systematic review. Applied Soft Computing **148**, 110915 (2023)
12. Wu, H., Zhang, Z., Wu, Q.: Exploring syntactic and semantic features for authorship attribution. Applied Soft Computing **111**, 107815 (2021)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_24