Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 192 (2021) 3922–3931

Procedia Computer Science

www.elsevier.com/locate/procedia

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Condition attributes, properties of decision rules, and discretisation: Analysis of relations and dependencies

Beata Zielosko[a,*], Urszula Stańczyk[b]

[a] *University of Silesia in Katowice, Będzińska 39, 41-200 Sosnowiec, Poland*
[b] *Department of Graphics, Computer Vision and Digital Systems, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Akademicka 2A, 44-100 Gliwice, Poland*

Abstract

When mining of input data is focused on rule induction, knowledge, discovered in exploration of existing patterns, is stored in combinations of certain conditions on attributes included in rule premises, leading to specific decisions. Through their properties, such as lengths, supports, cardinalities of rule sets, inferred rules characterise relations detected among variables. The paper presents research dedicated to analysis of these dependencies, considered in the context of various discretisation methods applied to the input data from stylometric domain. For induction of decision rules from data, Classical Rough Set Approach was employed. Next, based on rule properties, several factors were proposed and evaluated, reflecting characteristics of available condition attributes. They allowed to observe how variables and rule sets changed depending on applied discretisation algorithms.

Keywords: Discretisation; Decision rule; Rule length; Support; Attribute; Rough Sets

1. Introduction

Discretisation imposes granularity on the input space by construction of intervals, designed to represent ranges of values for considered attributes [1]. The process widens the scope of methods to be used for data mining, as some inducers operate only on nominal variables. Even when some classification system can work on both continuous and discrete features, categorical representation of values can bring such advantages as structure reduction that follows data reduction. When applied, typically discretisation constitutes a part of initial data preparation stage, after which data mining takes place [2]. It is also possible to firstly explore real-valued attributes, only to transform learned patterns [3]. As various discretisation methods exist, such reversal of steps allows to try out several algorithms, while avoiding the high computational costs of repeated knowledge discovery stage. Discretisation approaches can focus on many

* Corresponding author.
E-mail address: beata.zielosko@us.edu.pl

10.1016/j.procs.2021.09.167

criteria, and result in construction of different numbers of intervals for variables taken under consideration. Therefore, they all return some variants of input data that influence relationships and patterns present, and their exploration [4].

Knowledge discovered in data needs to be stored in some form [5]. Apart from being accessible for the process of labelling unknown examples, it is most advantageous when this form of representation provides also characterisation of attributes and detected dependencies, as it brings enhanced understanding. The transparency, such as shown by decision rules induced from input data, offers deep insight into existing relationships, and various rule parameters and rule based measures can be used for feature characterisation [6]. Within the experiments the rules were inferred by an exhaustive algorithm, dedicated to rough set processing of data.

Rough set theory enables description of sets by their approximations, and uses atoms of information corresponding to equivalence classes of objects that cannot be discerned based on values of known attributes [7]. This granularity is reflected in categorical variables that are required for classical rough set approach that was employed for mining stylometric data [8]. Exploration of stylistic characteristics of texts to the point of authorship attribution is considered as the most important task [9], and datasets used in research constituted examples of such problem. They were subjected to several discretisation algorithms, and from all versions of discrete datasets rule sets were inferred.

The methodology proposed in the framework of research allowed for in-depth analysis of relationships among sets of decision rules, induced from datasets discretised by various selected approaches, and attributes included in these rules. Most often rule lengths, supports and overall cardinalities of obtained rule sets are considered as important indicators of rule quality [10], and these properties were exploited in the experiments. They were used as a base for definition of new factors reflecting properties of individual attributes, relative to knowledge represented by sets of inferred rules. The proposed factors, when evaluated, led to obtaining several statistics for all considered attributes, and their analysis and comparison brought several observations on visible dependencies, and conclusions from results of combining different discretisation approaches applied to data.

The experiments performed consisted of several stages:

- preparation of input datasets;
- discretisation of data;
- induction of decision rules and their evaluation;
- creation of characteristics of attributes occurring in rule sets based on the proposed factors;
- construction of learning sets based on the obtained characteristics of attributes;
- induction of decision rules for modified learning sets;
- evaluation of obtained sets of rules and construction characteristics of attributes.

Comments on the properties of input features and data are presented in Section 2, along with explanation of various discretisation approaches. Description of properties of decision rules is given in Section 3, as are the proposed factors based on rules and dedicated to attribute characterisation. Observations on differences related to various discrete variants of datasets, induced rules, and attributes are included in Section 4, and final conclusions in Section 5.

2. Nature of Stylometry as the Application Domain and Input Datasets

Initial preparation of data included selection of: (i) authors and texts for authorship attribution tasks, (ii) characteristic features, (iii) discretisation approaches applied to data. This pre-processing stage of research was concluded with obtaining several selected variants of discrete input datasets, which were next subjected to data mining.

2.1. Input Features and Preparation of Data

Authorship attribution is considered as the most important task in the domain of stylometry [11]. It relies on construction of stylistic profiles for compared authors, reflecting their preferences with respect to linguistic elements, observable in many samples of writing [12]. Stylometric descriptors often refer to frequently employed function words, and punctuation marks as they are used habitually. Recognition of authorship can be treated as a classification task, where authors define classes and linguistic markers give characteristic features.

In the research for stylometric processing literary works of four renowned writers were chosen, and grouped into pairs: Edith Wharton and Mary Johnston, and Jack London and James Curwood. To obtain samples, the selected novels were divided in parts of comparable size. For all these text chunks there were calculated frequencies of occurrence for 22 words and 2 punctuation marks as follows: after, almost, any, around, before, but, by, during, how, never, on, same, such, that, then, there, though, until, whether, what, within, who, a comma, a semicolon.

The datasets prepared in this way contained 24 continuous characteristic features and 100 samples per author in a set (200 samples per set). In order to limit the number of factors influencing observations, balance of classes was ensured and classification was binary. To prepare data for rough set processing, discretisation was executed next.

2.2. Discretisation Algorithms

Right conduct regarding data preparation has an impact on the results obtained in the subsequent stages of data mining. Discretisation can be considered as an important element of data pre-processing step [2]. It transforms a set of continuous values into discrete ones, by partitioning ranges of numeric variables into a number of bins (sub-ranges) and treating each bin as a category or nominal value [4]. This division is based on selection of so-called *cut-points*.

To find intervals among ranges of attribute values *unsupervised* approaches only need to know their numbers, provided as an input parameter [13]. The main algorithms of this type are *equal width binning* and *equal frequency binning* [1]. For the former, values of the discretised attribute are sorted, then the minimum and maximum values are found, and finally all values in the range are divided into the number of equal width discrete intervals defined by a user. In the latter, after determining the minimum and maximum values of the discretised attribute, the range of values is divided into the specified number of intervals, such that each bin contains the same number of sorted values.

Supervised approaches consider information about class labels of objects when they investigate construction of intervals for attribute values. Known algorithms used are by *Fayyad and Irani* [14], and *Kononenko* [15]. Both methods use class entropy of constructed bins for evaluation of cut-points, and Minimum Description Length principle as a stopping criterion. Intervals are partitioned in top-down fashion, starting from one interval containing all values of the considered attribute, then the proposed cut-points are evaluated and the best one chosen for splitting the range of continuous values into sub-ranges. Discretisation is continued with each interval until a stopping criterion is achieved.

Let set S contain N instances and k decision classes $C_1, \ldots, C_k$. For $P(C_i, S)$ denoting the proportion of class C_i instances in S, class entropy $Ent(S)$ is defined as follows:

$$Ent(S) = -\sum_{i=1}^{k} P(C_i, S)\log(P(C_i, S)). \tag{1}$$

For binary discretisation of a continous attribute A, a candidate cut-point T splits set S into two subsets, S_1 and S_2, where $S1 \subset S$ contains instances with attribute values $\leq T$ and $S_2 = S \setminus S_1$. Entropy for cut-point T is obtained by:

$$Ent(A, T; S) = \frac{|S_1|}{|S|}Ent(S_1) + \frac{|S_2|}{|S|}Ent(S_2). \tag{2}$$

The selection of cut-point T_{opt} is made by testing all candidates. For the optimal cut-point T_{opt} class information entropy $Ent(A, T_{opt}; S)$ is minimal. Evaluation of cut-points is executed recursively until the stop criterion is met.

For Fayyad and Irani method, the stopping criterion, given by inequality Eq. (3), is based on information gain resulting from the cut-point T, $Gain(A, T; S) = Ent(S) - E(A, T; S)$.

$$Gain(A, T; S) > \frac{\log_2(N-1)}{N} + \frac{\log_2(3^k - 2) - [k \cdot Ent(S) - k_1 \cdot Ent(S_1) - k_2 \cdot Ent(S_2)]}{N}. \tag{3}$$

In the case of Kononenko algorithm, let N be the number of training instances, N_{C_i} denote the number of training instances from the class C_i, N_{A_x} specify the number of instances with x-th value of the given attribute, $N_{C_iA_y}$ the number of instances from class C_i with y-th value of the given attribute, and N_T the number of possible cut-points. The stopping criterion requires that the inequality (4) becomes true:

$$\log\binom{N}{N_{C_1}\dots N_{C_k}} + \log\binom{N+k-1}{k-1} > \sum_j \log\binom{N_{A_j}}{N_{C_1A_j}\dots N_{C_kA_j}} + \sum_j \binom{N_{A_j}+k-1}{k-1} + \log N_T. \quad (4)$$

In majority of research, once the decision with respect to application of discretisation is made, all attributes are transformed, and to all variables the same algorithm is applied. This popular practice is not necessarily the most advantageous, as sometimes processing of only a part of attributes, instead of all, can bring better results [16]. The supervised methods are often considered as more efficient and delivering better results then unsupervised, in particular equal width binning is frequently criticised as inflexible to data distributions [17]. However, combinations of methods can allow for smaller reduction of information leading to enhanced predictions [18, 19].

2.3. Discrete Datasets

The choice of some particular discretisation algorithm can significantly influence data mining and knowledge discovery process [20]. In the research several variants of discrete input data were obtained and compared in the perspective of properties of rule sets induced. These datasets were as follows:

- dsF—supervised discretisation with Fayyad and Irani method applied to all attributes;
- dsK—supervised discretisation with Kononenko algorithm for all attributes;
- duf*i*—unsupervised equal frequency binning, with *i* giving the number of intervals, the same for all attributes;
- duw*i*—unsupervised equal width binning, with *i* giving the number of intervals, the same for all attributes;
- factor-based—combination of specific supervised and unsupervised approaches, selected individually for attributes depending on observed values for the proposed factors referring to rule sets induced from data.

The input parameter i used for unsupervised approaches was from the set $\{2, 3, 4, 5\}$. The upper limit considered was chosen by an analysis of the maximum numbers of bins found for attributes by supervised algorithms. They ranged from 1, which means a single interval representing all values of some attribute, to the maximum of 4 intervals. Such 1-bin variables brought no informative content in discrete space [6], and were in fact disregarded within knowledge discovery process that was the next stage of experiments.

3. Knowledge Represented by Decision Rules

Decision rules are popular and very useful form of knowledge representation [21], since their notation is similar to the way how a human writes knowledge. Simplicity of understanding and interpretation by domain experts are considered their main advantages, and the reason for frequent application in many areas connected with data mining and knowledge discovery. Due to this popularity, a variety of approaches and algorithms for decision rules construction exist [22]. In the reported research, decision rules were induced in the framework of rough sets theory [7].

3.1. Properties of Decision Rules

In rough set universe U, objects are described by data stored in a tabular form, known as a *decision table*: $S = (U, A \bigcup \{d\})$. $A = \{a_1, \dots, a_m\}$ is a nonempty, finite set of condition attributes, i.e., $a_i : U \to V_a$, where V_a is the set of values of attribute a_i, and d is a distinguished attribute $d \notin A$ called a decision. *Decision rules* induced from the table are expressions presented in a form $(a_{i_1} = v_1) \wedge \dots \wedge (a_{i_k} = v_k) \to d = v_d$, where $1 \le i_1 < \dots < i_k \le m, v_i \in V_{a_i}$, $1 \le v_d \le |V_d|$. In the research they were inferred by the exhaustive algorithm, which constructs all decision rules with minimal number of pairs *attribute = value* in the premise part of the rule [23].

The aim of discovery oriented induction is to find patterns or regularities hidden in the data set, which are interesting and useful for users. Once decision rules are induced [24], they can also be analysed from the point of view of knowledge representation. Their sets can be studied through the perspective of their properties and measures used for evaluation of their quality [25]. Typically the considered rule characteristics refer to:

- cardinality of rule sets;
- lengths of rules, giving the number of descriptors (conditions) in premise part of the rule;
- supports of rules, which is the number of objects from the training dataset covered by the rule.

Rules with fewer conditions are considered as more advantageous. They are more general, possess better descriptive properties, and are simpler for interpretation. Heuristics for optimisation of decision rule induction often focus on this aspect and return only subsets of rules with specific lengths [26]. High support of rules indicates that captured patterns were observed in many training objects, which is naturally preferred over rare cases, as it speaks of rule strength.

3.2. Induced Decision Rules

In the research Rough Set Exploration System was employed [27] for induction of decision rules. For sets of rules induced from data discretised by selected algorithms, Table 1 presents evaluation measures such as the number of included rules, average and maximum values of length and support (minima were equal 1 for all rule sets).

Table 1. Properties of rule sets induced from data discretised by various algorithms, for Female and Male writer datasets

Female writer datasets						Male writer datasets					
rule	number	length		support		rule	number	length		support	
set	of rules	average	maximum	average	maximum	set	of rules	average	maximim	average	maximum
dsF	4121	4.8	9	6.6	88	dsF	15283	5.1	9	5.9	79
dsK	10190	5.3	10	5.4	88	dsK	20815	5.1	10	5.5	75
duf2	103645	5.7	11	3.7	92	duf2	138910	5.6	10	3.6	56
duf3	122527	4.5	8	2.1	66	duf3	135696	4.4	7	2.0	67
duf4	81723	3.8	7	1.8	50	duf4	96327	3.7	7	1.8	50
duf5	68994	3.5	6	1.7	40	duf5	76240	3.4	7	1.7	40
duw2	2094	5.5	11	6.5	86	duw2	1509	4.6	10	7.5	78
duw3	26025	5.6	10	3.6	75	duw3	32447	5.7	12	3.3	66
duw4	46480	5.0	9	2.6	58	duw4	47574	5.0	10	2.9	49
duw5	67054	4.5	9	2.1	54	duw5	79561	4.6	9	2.3	51

It can be observed that various discretisation methods applied to the datasets resulted in highly varied characteristics of rule sets induced from them, in all aspects. Different numbers of rules were inferred, with varying averages of lengths and supports. Supervised discretisation approaches not always resulted in the best rule sets. The lowest average rule length was related to lowest maximal length, and highest average supports with rule sets of lowest cardinalities.

These characteristics can be used not only for comparisons of rule sets, but also to describe individual attributes used as conditions in rule premises. With that aim in the research several factors were defined, based on rules and focused in attributes, as explained in the next section.

3.3. Characterisation of Attributes by Rules and Rule Sets

Given a set of decision rules S_{Rls}, $card(S_{Rls})$ denotes the cardinality of the rule set, that is the number of rules in the set. For an attribute a, $RLS(S_{Rls}, a)$ is the subset of rules from the set S_{Rls} that include conditions on a attribute. It can be argued that if an attribute occurs in many rules, then it is considered as important for rule induction process. Such factor was defined by Eg. (5). Its modification, given by Eq. (6), enabled to calculate the average length of rules including the attribute a. Taking into account the preference of shorter rules led to Eg. (7), where rules were weighted by their lengths. Eg. (8) brought information on average support of rules in which the attribute occurred.

$$(NoR) \qquad W_N(S_{Rls}, a) = \frac{card(RLS(S_{Rls}, a))}{card(S_{Rls})}, \tag{5}$$

$$(AvgL)\ W_{AL}(S_{Rls}, a) = \sum_{i=l_{min}}^{l_{max}} \frac{card(RLS(S_{Rls}, a, i))}{card(S_{Rls})} \cdot i, \tag{6}$$

$$(Wl) \quad W_L(S_{Rls}, a) = \sum_{i=l_{min}}^{l_{max}} \frac{card(RLS(S_{Rls}, a, i))}{card(S_{Rls}) \cdot i}, \tag{7}$$

$$(AvgS)\ W_S(S_{Rls}, a) = \sum_{i=s_{min}}^{s_{max}} \frac{card(RLS(S_{Rls}, a, i))}{card(S_{Rls})} \cdot i, \tag{8}$$

In calculation of all these factors the total number of rules in the considered rule set was taken into account, that is occurrence in rules was always treated as relative to all rules in the analysed set. In the rest of the paper the factors will be referred to by their labels given on the left in the above equations. With these definitions, for *NoR* highest values were preferred, and the same statement was true for weighted lengths *Wl* and averaged support *AvgS*, whereas the smallest values for averaged lengths *AvgL* were considered as most advantageous.

4. From Datasets to Rule Sets then to Attributes

For all variants of discrete input datasets and rule sets induced from them, statistics were obtained for attributes, exploiting the previously defined factors built around rules and their properties. These individual characteristics were analysed in relation to discretisation methods used, which led to modifications of the training sets. From the new sets again rules were inferred, and again statistics calculated and compared. In all computations representation included 4 fractional digits. In the presented tables they were rounded to at most 2, due to space limitation. Zeroes indicate absence of an attribute in a considered rule set. In all tables preferred or improved values were indicated with colours.

4.1. Statistics Obtained for Attributes

NoR, defined by Eq. (5), was the first statistics considered. It corresponded to the percentage of rules in a rule set studied, which contained a given attribute. The characteristics, displayed by Table 2, show that for female writer dataset the highest number of variables (8) indicated equal width binning with just 2 bins as leading to the highest values of *NoR* factor, the remaining ones pointed to both supervised methods, equal frequency binning with 2 bins, and again equal width binning but with 3 bins this time. For male writers the distribution of selections among discretisation methods was closer, for the highest number of variables (7) it was equal frequency binning with 2 bins, for the rest Fayyad and Irani supervised discretisation, and equal width binning with either 2 or 3 bins. Overall, methods leading to smaller numbers of intervals constructed were mostly selected, but from both supervised and unsupervised approaches.

Table 3 shows statistics based on the second factor *AvgL* (Eg. (6)), calculating the average length of rules containing conditions on a attribute, in relation to the total number of decision rules in a set. In the case of this factor, smallest values were preferred. With only few exceptions, for both female and male datasets, the vast majority of attributes distinguished equal frequency binning with 5 bins as the method leading to shortest rules. This observation brought an interesting contrast when compared with the content of Table 4, which reflects *Wl* factor (Eg. (7)) that was also based on rule lengths. However, in this case the preference for shorter rules was incorporated directly into calculations, through weighting rules by their lengths. For female writers *Wl* factor pointed mostly to Fayyad and Irani algorithm of supervised discretisation, and for male writers to Fayyad and equal width binning with 2 bins. In both cases for most variables smaller numbers of bins were constructed than in algorithms pointed by *AvgL* factor.

Table 5 presents characteristics based on the last of considered factors, *AvgS* (Eg. (8)), which took into account supports of rules with conditions on a given attribute. As rules with high supports are preferred as stronger, highest values were distinguished. From all discretisation methods the ones that stood out as leading to the best statistics for both datasets were Fayyad and Irani supervised algorithm, and equal width binning with 2 bins. Actually, when statistics for all considered factors were compared, putting aside the characterisation by *AvL* factor, the rest pointed mostly to these two discretisation procedures for both female and male writer datasets, as returning preferred values.

Table 2. Statistics obtained for attributes based on *NoR* factor (see Eg. (5)), presenting the percentage of rules in which an attribute is included

rule set	attr0	attr1	attr2	attr3	attr4	attr5	attr6	attr7	attr8	attr9	attr10	attr11	attr12	attr13	attr14	attr15	attr16	attr17	attr18	attr19	attr20	attr21	attr22	attr23
Female writer datasets																								
dsF	4.1	29.7	30.1	35.0	29.0	4.5	31.8	24.8	34.5	38.3	28.1	18.9	39.0	26.7	0.0	30.1	0.0	13.6	0.0	7.1	19.7	0.0	27.3	9.2
dsK	3.1	24.0	30.4	31.2	30.6	3.7	28.8	25.8	34.5	38.4	27.2	19.7	36.7	26.8	23.9	29.5	18.0	10.4	0.0	9.0	15.7	36.2	23.5	5.8
duf2	28.6	12.7	24.4	28.9	26.2	29.2	24.2	31.3	26.6	29.1	28.8	31.7	29.1	23.5	16.6	30.5	30.3	11.0	28.4	16.6	15.7	28.3	15.7	2.6
duf3	22.5	11.5	17.8	21.6	21.5	23.9	19.6	21.7	20.5	22.5	23.0	22.5	20.9	18.4	11.5	23.4	22.8	9.9	20.3	10.6	11.5	21.0	17.8	9.5
duf4	20.6	10.5	15.7	19.7	18.9	20.0	16.4	19.6	18.1	19.1	18.9	20.5	17.9	15.6	9.0	20.5	19.5	9.5	16.6	8.1	10.5	18.1	14.0	4.4
duf5	17.7	10.1	13.6	18.1	17.4	17.3	15.3	17.3	16.3	17.1	17.2	18.5	16.3	14.3	8.0	19.1	18.4	8.8	15.0	6.6	9.3	15.8	12.9	5.2
duw2	43.6	37.1	38.1	14.0	33.4	21.9	27.7	23.7	14.5	33.1	13.2	32.3	27.6	29.6	3.1	30.7	24.3	4.1	28.9	13.3	2.0	3.0	37.8	11.1
duw3	33.9	22.7	29.8	16.9	31.3	27.7	29.4	29.0	27.7	33.8	34.4	32.5	27.4	21.1	6.3	29.8	23.9	7.5	25.6	5.7	5.4	12.7	28.2	14.3
duw4	28.4	18.8	24.4	24.9	25.9	26.7	27.9	23.7	24.6	27.3	24.3	29.4	20.4	20.7	4.4	29.3	28.6	6.4	19.6	15.7	4.7	11.6	23.5	7.0
duw5	23.3	13.3	19.0	22.9	22.9	25.1	21.4	24.0	23.0	21.9	24.7	24.7	22.3	15.8	2.9	25.1	22.8	10.1	22.9	12.3	10.2	20.1	16.5	4.1
Male writer datasets																								
rule set	attr0	attr1	attr2	attr3	attr4	attr5	attr6	attr7	attr8	attr9	attr10	attr11	attr12	attr13	attr14	attr15	attr16	attr17	attr18	attr19	attr20	attr21	attr22	attr23
dsF	29.2	26.3	16.4	20.8	22.3	30.5	26.8	23.7	26.0	27.2	17.6	0.0	30.1	27.7	0.0	20.5	20.0	27.9	15.3	22.5	28.7	16.1	20.9	10.4
dsK	27.6	25.4	27.1	18.8	21.5	29.5	24.0	22.7	24.6	24.9	15.6	15.2	28.6	27.1	0.0	20.3	19.5	25.2	15.7	20.8	27.0	14.5	19.8	9.8
duf2	21.0	19.9	20.5	17.8	26.7	28.6	18.2	25.1	26.0	26.0	28.5	29.1	27.9	24.4	22.0	30.9	27.2	19.9	24.6	15.3	25.1	24.9	27.0	7.9
duf3	19.6	18.3	16.4	17.4	20.4	23.0	15.2	19.0	20.6	17.9	22.7	23.0	21.3	19.1	15.7	22.1	16.3	16.3	18.1	10.2	18.8	15.9	20.4	9.9
duf4	15.8	14.6	14.5	14.2	18.5	19.4	14.2	16.8	16.8	16.2	19.3	19.4	18.9	16.3	12.8	18.0	15.5	13.6	16.8	8.4	15.7	13.2	17.5	8.4
duf5	14.6	13.2	13.5	14.0	16.7	17.5	13.5	16.1	15.8	13.9	18.0	17.7	16.8	14.2	11.0	17.2	13.5	12.8	14.7	7.7	14.1	10.5	15.7	5.5
duw2	38.2	33.6	29.2	29.1	30.7	18.2	15.5	21.9	29.1	32.7	23.3	24.3	8.7	9.1	1.9	11.9	9.6	18.4	12.7	10.9	13.0	3.3	10.3	27.0
duw3	29.4	29.0	28.4	23.8	31.0	28.7	27.6	26.7	33.8	27.6	33.8	27.6	19.3	17.2	8.2	21.6	19.6	21.2	18.2	19.4	20.2	9.6	26.8	21.2
duw4	23.7	22.3	22.6	23.8	27.6	28.3	19.9	21.2	26.3	22.4	26.9	26.0	20.5	15.9	5.4	29.3	18.4	18.9	18.3	16.5	16.4	8.0	24.1	12.9
duw5	20.4	17.7	18.6	18.9	22.5	22.7	18.8	20.1	22.3	22.2	24.9	20.9	20.5	18.0	15.9	25.4	13.7	15.0	16.4	12.7	19.2	17.5	22.5	8.5

Table 3. Statistics obtained for attributes based on *AvgL* factor (see Eg. (6)), giving average length of rules including a given attribute, relative to the total number of rules

rule set	attr0	attr1	attr2	attr3	attr4	attr5	attr6	attr7	attr8	attr9	attr10	attr11	attr12	attr13	attr14	attr15	attr16	attr17	attr18	attr19	attr20	attr21	attr22	attr23
Female writer datasets																								
dsF	0.20	1.42	1.52	1.80	1.52	0.24	1.61	1.32	1.76	2.01	1.51	1.02	2.02	1.43	0.0	1.58	0.0	0.64	0.0	0.42	0.97	0.0	1.38	0.36
dsK	0.17	1.24	1.71	1.75	1.76	0.22	1.58	1.50	1.92	2.19	1.58	1.17	2.06	1.58	1.43	1.69	1.06	0.56	0.0	0.57	0.85	2.05	1.27	0.24
duf2	1.68	0.72	1.41	1.69	1.53	1.71	1.40	1.84	1.55	1.70	1.69	1.86	1.71	1.35	1.00	1.79	1.78	0.63	1.66	1.03	0.92	1.65	0.90	0.13
duf3	1.02	0.51	0.80	0.98	0.98	1.09	0.88	0.98	0.92	1.02	1.05	1.03	0.95	0.83	0.57	1.06	1.04	0.48	0.94	0.52	0.55	0.97	0.80	0.44
duf4	0.81	0.40	0.61	0.77	0.74	0.78	0.63	0.76	0.70	0.74	0.73	0.80	0.71	0.61	0.38	0.80	0.76	0.40	0.67	0.35	0.44	0.72	0.53	0.17
duf5	0.63	0.34	0.47	0.64	0.61	0.61	0.53	0.61	0.57	0.60	0.61	0.65	0.59	0.50	0.32	0.67	0.65	0.35	0.56	0.26	0.36	0.58	0.45	0.18
duw2	2.83	2.39	2.41	1.03	2.23	1.49	1.52	1.52	0.70	2.06	0.98	1.89	1.86	1.87	0.23	1.80	1.51	0.22	1.86	0.95	0.11	0.15	2.39	0.53
duw3	1.95	1.35	1.78	1.04	1.83	1.60	1.72	1.75	1.60	1.97	2.04	1.96	1.65	1.29	0.43	1.76	1.46	0.44	1.54	0.38	0.33	0.76	1.60	0.78
duw4	1.48	0.97	1.27	1.33	1.31	1.40	1.44	1.24	1.28	1.40	1.26	1.53	1.07	1.06	0.28	1.53	1.51	0.36	1.07	0.85	0.28	0.64	1.21	0.34
duw5	1.09	0.60	0.89	1.11	1.06	1.19	0.99	1.15	1.07	1.01	1.17	1.15	1.03	0.74	0.15	1.16	1.07	0.54	1.08	0.62	0.52	0.98	0.75	0.17
Male writer datasets																								
rule set	attr0	attr1	attr2	attr3	attr4	attr5	attr6	attr7	attr8	attr9	attr10	attr11	attr12	attr13	attr14	attr15	attr16	attr17	attr18	attr19	attr20	attr21	attr22	attr23
dsF	1.58	1.39	0.96	1.08	1.14	1.65	1.40	1.24	1.40	1.46	0.97	0.0	1.61	1.48	0.0	1.09	1.09	1.44	0.81	1.20	1.54	0.86	1.09	0.49
dsK	1.48	1.33	1.41	0.98	1.11	1.58	1.24	1.19	1.31	1.33	0.85	0.82	1.52	1.45	0.0	1.09	1.06	1.29	0.83	1.11	1.44	0.77	1.04	0.46
duf2	1.20	1.14	1.18	1.02	1.55	1.67	1.03	1.45	1.52	1.51	1.66	1.70	1.62	1.42	1.28	1.81	1.58	1.13	1.43	0.90	1.45	1.45	1.57	0.43
duf3	0.88	0.81	0.73	0.77	0.91	1.03	0.67	0.84	0.92	0.79	1.03	1.03	0.95	0.85	0.72	0.99	0.72	0.72	0.81	0.48	0.85	0.75	0.91	0.45
duf4	0.60	0.55	0.55	0.54	0.71	0.75	0.53	0.64	0.65	0.62	0.75	0.75	0.73	0.63	0.51	0.69	0.60	0.52	0.65	0.35	0.62	0.54	0.67	0.32
duf5	0.50	0.45	0.46	0.47	0.58	0.60	0.46	0.56	0.54	0.48	0.62	0.62	0.59	0.50	0.41	0.59	0.47	0.44	0.51	0.30	0.51	0.40	0.54	0.19
duw2	1.92	1.83	1.44	1.58	1.67	1.03	0.72	1.25	1.48	1.79	1.14	1.33	0.49	0.52	0.09	0.55	0.60	0.96	0.70	0.60	0.73	0.17	0.55	1.25
duw3	1.78	1.77	1.71	1.36	1.87	1.75	1.61	1.66	2.03	1.62	2.03	1.70	1.26	1.09	0.53	1.36	1.29	1.31	1.05	1.22	1.28	0.60	1.65	1.30
duw4	1.21	1.18	1.14	1.22	1.45	1.46	0.99	1.16	1.33	1.15	1.40	1.37	1.13	0.88	0.31	1.58	0.98	0.94	0.96	0.91	0.93	0.47	1.29	0.64
duw5	0.94	0.82	0.87	0.87	1.07	1.07	0.88	0.97	1.02	1.05	1.17	1.00	1.02	0.88	0.78	1.23	0.67	0.69	0.79	0.64	0.93	0.88	1.07	0.38

Based on observations drawn from evaluation of factors for all rule sets and all attributes considered, modified versions of discrete input datasets were constructed next. For each attribute such discretisation method was selected that led to obtaining the most preferred value for each considered factor. With preferences indicating both supervised and unsupervised algorithms, the resulting datasets constituted examples of combinations of several different discretisation approaches applied to data in a single dataset, which is not a standard procedure.

Table 4. Statistics obtained for attributes based on *Wl* factor (see Eg. (7)), with weighting rules including an attribute by their lengths, relative to the total number of rules, multiplied by 10

rule set	attr0	attr1	attr2	attr3	attr4	attr5	attr6	attr7	attr8	attr9	attr10	attr11	attr12	attr13	attr14	attr15	attr16	attr17	attr18	attr19	attr20	attr21	attr22	attr23
Female writer datasets																								
dsF	0.26	1.94	1.86	2.08	1.72	0.27	1.93	1.44	2.06	2.24	1.61	1.09	2.30	1.53	0.00	1.77	0.00	0.91	0.00	0.36	1.25	0.00	1.64	0.74
dsK	0.06	0.50	0.58	0.59	0.56	0.07	0.56	0.47	0.65	0.71	0.49	0.35	0.69	0.48	0.42	0.54	0.33	0.21	0.00	0.15	0.32	0.67	0.46	0.15
duf2	0.50	0.24	0.43	0.50	0.46	0.51	0.43	0.55	0.47	0.51	0.50	0.55	0.51	0.42	0.28	0.53	0.53	0.20	0.50	0.28	0.28	0.50	0.28	0.05
duf3	0.51	0.27	0.40	0.49	0.48	0.54	0.44	0.49	0.47	0.51	0.52	0.51	0.47	0.42	0.24	0.53	0.51	0.21	0.45	0.22	0.25	0.46	0.41	0.21
duf4	0.54	0.29	0.42	0.52	0.50	0.52	0.44	0.52	0.48	0.50	0.50	0.54	0.47	0.42	0.22	0.54	0.52	0.23	0.43	0.20	0.26	0.47	0.38	0.12
duf5	0.52	0.30	0.40	0.53	0.51	0.51	0.45	0.51	0.48	0.51	0.50	0.54	0.47	0.43	0.21	0.56	0.54	0.23	0.42	0.18	0.25	0.44	0.38	0.15
duw2	0.76	0.65	0.69	0.22	0.56	0.37	0.57	0.43	0.34	0.60	0.21	0.61	0.46	0.53	0.06	0.59	0.45	0.09	0.51	0.20	0.04	0.08	0.67	0.28
duw3	0.63	0.41	0.53	0.29	0.57	0.50	0.53	0.51	0.51	0.61	0.61	0.57	0.49	0.37	0.10	0.54	0.42	0.13	0.45	0.09	0.10	0.23	0.53	0.28
duw4	0.57	0.39	0.49	0.49	0.54	0.53	0.56	0.48	0.50	0.56	0.49	0.59	0.41	0.42	0.07	0.59	0.57	0.12	0.38	0.30	0.09	0.23	0.48	0.16
duw5	0.52	0.31	0.43	0.49	0.52	0.55	0.48	0.52	0.52	0.50	0.54	0.55	0.50	0.36	0.06	0.57	0.50	0.20	0.50	0.26	0.21	0.43	0.39	0.11
Male writer datasets																								
rule set	attr0	attr1	attr2	attr3	attr4	attr5	attr6	attr7	attr8	attr9	attr10	attr11	attr12	attr13	attr14	attr15	attr16	attr17	attr18	attr19	attr20	attr21	attr22	attr23
dsF	0.57	0.52	0.29	0.42	0.45	0.59	0.54	0.48	0.50	0.53	0.34	0.00	0.59	0.54	0.00	0.41	0.39	0.56	0.31	0.44	0.56	0.32	0.42	0.24
dsK	0.53	0.50	0.54	0.38	0.43	0.57	0.48	0.45	0.48	0.49	0.30	0.30	0.56	0.53	0.00	0.40	0.38	0.51	0.32	0.41	0.53	0.29	0.39	0.22
duf2	0.38	0.36	0.37	0.32	0.47	0.50	0.33	0.45	0.46	0.46	0.50	0.51	0.49	0.43	0.39	0.54	0.48	0.36	0.44	0.27	0.44	0.44	0.48	0.15
duf3	0.45	0.42	0.38	0.40	0.47	0.52	0.35	0.44	0.47	0.41	0.52	0.52	0.49	0.44	0.35	0.50	0.38	0.38	0.41	0.22	0.42	0.35	0.47	0.22
duf4	0.43	0.40	0.39	0.39	0.49	0.52	0.39	0.45	0.45	0.44	0.51	0.51	0.50	0.43	0.33	0.48	0.41	0.37	0.45	0.21	0.41	0.34	0.47	0.23
duf5	0.44	0.40	0.41	0.42	0.50	0.52	0.41	0.48	0.47	0.41	0.53	0.53	0.49	0.42	0.31	0.52	0.40	0.39	0.44	0.21	0.40	0.29	0.47	0.17
duw2	0.82	0.70	0.66	0.60	0.64	0.36	0.38	0.43	0.64	0.68	0.53	0.50	0.19	0.20	0.06	0.29	0.17	0.41	0.28	0.24	0.28	0.08	0.23	0.65
duw3	0.52	0.51	0.50	0.44	0.55	0.51	0.50	0.46	0.60	0.50	0.60	0.48	0.31	0.29	0.14	0.37	0.31	0.38	0.34	0.33	0.34	0.17	0.46	0.37
duw4	0.49	0.45	0.48	0.49	0.56	0.58	0.42	0.41	0.55	0.46	0.55	0.52	0.39	0.31	0.10	0.58	0.36	0.40	0.37	0.32	0.31	0.15	0.48	0.28
duw5	0.46	0.40	0.42	0.43	0.50	0.51	0.42	0.44	0.51	0.49	0.55	0.46	0.43	0.39	0.34	0.55	0.29	0.35	0.36	0.26	0.41	0.37	0.49	0.20

Table 5. Statistics obtained for attributes based on *AvgS* factor (see Eg. (8)), giving average support of rules including a given attribute, relative to the total number of rules

rule set	attr0	attr1	attr2	attr3	attr4	attr5	attr6	attr7	attr8	attr9	attr10	attr11	attr12	attr13	attr14	attr15	attr16	attr17	attr18	attr19	attr20	attr21	attr22	attr23
Female writer datasets																								
dsF	0.40	1.88	1.39	1.89	1.58	0.51	1.74	1.32	2.08	2.04	1.44	0.91	2.06	1.33	0.00	1.38	0.00	1.18	0.00	0.28	1.44	0.00	1.66	0.66
dsK	0.23	1.17	1.25	1.42	1.36	0.36	1.49	1.20	1.79	1.72	1.17	0.83	1.68	1.16	1.14	1.20	0.71	0.77	0.00	0.36	1.03	1.61	1.23	0.35
duf2	0.90	0.53	0.82	0.91	0.85	0.92	0.84	0.98	0.88	0.95	0.90	0.98	0.91	0.79	0.56	0.96	0.93	0.51	0.94	0.47	0.59	0.91	0.57	0.17
duf3	0.42	0.26	0.35	0.41	0.41	0.44	0.38	0.41	0.40	0.43	0.43	0.42	0.40	0.36	0.24	0.44	0.43	0.26	0.40	0.20	0.27	0.41	0.35	0.18
duf4	0.34	0.20	0.27	0.33	0.32	0.33	0.28	0.33	0.31	0.32	0.32	0.33	0.31	0.27	0.17	0.34	0.33	0.21	0.29	0.14	0.22	0.32	0.26	0.10
duf5	0.27	0.18	0.22	0.28	0.27	0.27	0.25	0.27	0.26	0.27	0.26	0.28	0.26	0.23	0.14	0.29	0.28	0.19	0.24	0.11	0.19	0.26	0.21	0.09
duw2	2.40	2.56	2.35	0.76	1.98	1.59	1.83	1.04	0.80	2.20	0.89	2.23	1.52	1.76	0.09	1.94	1.70	0.29	1.57	0.77	0.12	0.17	2.08	1.50
duw3	1.04	0.77	0.98	0.55	1.17	0.96	1.04	1.07	0.88	1.26	1.13	1.12	1.02	0.90	0.30	1.02	0.82	0.31	0.98	0.19	0.25	0.40	1.00	0.47
duw4	0.63	0.47	0.62	0.68	0.62	0.67	0.68	0.66	0.59	0.68	0.63	0.72	0.58	0.64	0.17	0.69	0.73	0.17	0.52	0.38	0.11	0.29	0.56	0.17
duw5	0.44	0.29	0.38	0.48	0.45	0.49	0.43	0.49	0.44	0.45	0.48	0.49	0.45	0.38	0.08	0.47	0.46	0.22	0.48	0.25	0.23	0.41	0.34	0.08
Male writer datasets																								
rule set	attr0	attr1	attr2	attr3	attr4	attr5	attr6	attr7	attr8	attr9	attr10	attr11	attr12	attr13	attr14	attr15	attr16	attr17	attr18	attr19	attr20	attr21	attr22	attr23
dsF	1.58	1.31	0.94	1.23	1.11	1.53	1.65	1.19	1.39	1.41	0.86	0.00	1.54	1.36	0.00	0.97	1.22	1.62	0.90	1.09	1.42	0.82	1.10	0.45
dsK	1.31	1.08	1.23	0.97	0.94	1.23	1.37	0.97	1.14	1.13	0.69	0.68	1.24	1.15	0.00	0.82	1.06	1.31	0.81	0.90	1.19	0.60	0.92	0.33
duf2	0.73	0.70	0.70	0.60	0.81	0.85	0.64	0.78	0.78	0.81	0.86	0.87	0.86	0.74	0.70	0.92	0.85	0.70	0.78	0.53	0.78	0.77	0.84	0.35
duf3	0.37	0.35	0.33	0.34	0.38	0.42	0.31	0.37	0.38	0.35	0.41	0.41	0.40	0.36	0.31	0.40	0.33	0.33	0.35	0.23	0.35	0.30	0.38	0.17
duf4	0.27	0.25	0.25	0.25	0.31	0.32	0.26	0.29	0.28	0.28	0.31	0.31	0.31	0.28	0.22	0.30	0.27	0.24	0.28	0.17	0.26	0.23	0.29	0.14
duf5	0.23	0.21	0.21	0.22	0.26	0.27	0.22	0.26	0.24	0.23	0.27	0.27	0.26	0.24	0.18	0.26	0.23	0.21	0.24	0.14	0.23	0.17	0.24	0.10
duw2	3.06	2.96	2.49	2.28	1.71	1.27	1.40	1.50	1.86	2.23	1.53	1.95	1.19	0.42	0.11	0.83	0.46	1.18	0.96	0.49	0.80	0.19	0.53	3.33
duw3	0.92	0.93	0.87	0.93	0.92	0.92	0.90	0.88	1.00	0.86	1.06	0.84	0.62	0.58	0.28	0.71	0.68	0.65	0.58	0.58	0.64	0.33	0.85	0.51
duw4	0.65	0.62	0.63	0.63	0.68	0.73	0.56	0.55	0.68	0.62	0.71	0.73	0.58	0.47	0.13	0.76	0.62	0.52	0.46	0.49	0.47	0.24	0.68	0.36
duw5	0.44	0.41	0.41	0.39	0.47	0.48	0.40	0.46	0.44	0.49	0.50	0.47	0.46	0.42	0.33	0.54	0.40	0.34	0.35	0.32	0.45	0.37	0.49	0.18

4.2. Properties of Rule Sets Inferred from Modified Training Sets

From the modified training sets new sets of decision rules were induced, with corresponding properties and measures listed in Table 6. From all approaches tested, focus on *AvgL* factor in fact led to obtaining rule sets with the decreased average rule lengths. Other measures (cardinalities and average supports) were not better than the best previously obtained, but *NoR* factor caused the highest cardinalities of rule sets, confirming the aim at inclusion in high numbers of rules for attributes. And similarly to observations referring to Table 1, also in this case the maximum values of average support existed for datasets with the smallest numbers of rules among those considered.

Table 6. Properties of rule sets inferred from modified datasets

	Female writer datasets					Male writer datasets					
	number	length		support		number	length			support	
factor	of rules	average	maximum	average	maximum	of rules	min	average	maximum	average	maximum
NoR	55723	6.0	11	4.1	73	94956	1	5.7	11	3.8	64
AvgL	33291	3.4	7	1.8	38	44251	1	3.3	6	1.7	40
Wl	31494	5.6	10	4.1	88	69602	1	5.3	10	3.6	64
AvgS	15117	6.1	11	5.2	86	22411	2	5.4	11	5.9	79

Based on the newly inferred rule sets again statistics for attributes were obtained, given in Table 7. The prefixes F- and M- indicate respectively female or male writer dataset, and coloured cells distinguish cases of improvement. For both datasets some enhanced results could be observed, not always for the same attributes or factors. Part of the improvement was the consequence of applying unsupervised discretisation to variables for which supervised discretisation returned single bins, practically eliminating them from all considerations. This additional processing enabled access to the information brought by such variables at the data exploration stage, that was denied by standard supervised algorithms. Even if entropy-based measures evaluated their informative content as negligible, it was higher than zero, and brought benefits with extra transformations, as shown in the improved statistics.

Table 7. Statistics obtained for attributes based on considered factors (see Sect.3.3), calculated with relation to modified versions of both datasets

factor	attr0	attr1	attr2	attr3	attr4	attr5	attr6	attr7	attr8	attr9	attr10	attr11	attr12	attr13	attr14	attr15	attr16	attr17	attr18	attr19	attr20	attr21	attr22	attr23
F-*NoR*	30.3	22.8	25.2	26.2	24.1	34.0	24.1	34.2	29.7	33.0	30.2	30.9	28.1	18.0	22.5	20.8	35.0	8.6	15.4	21.1	11.3	32.6	29.7	10.5
M-*NoR*	21.5	21.5	20.6	21.2	26.9	30.5	20.6	22.4	25.4	18.9	28.9	30.2	25.4	25.6	20.6	31.0	28.2	21.1	25.3	19.4	26.0	24.3	28.6	8.5
F-*AvgL*	0.18	0.46	0.59	0.73	0.72	0.12	0.64	0.70	0.65	0.72	0.72	0.76	0.66	0.59	0.07	0.77	0.74	0.16	0.63	0.30	0.09	0.16	0.51	0.15
M-*AvgL*	0.56	0.53	0.52	0.55	0.64	0.68	0.51	0.62	0.61	0.54	0.69	0.68	0.11	0.57	0.05	0.11	0.53	0.50	0.56	0.33	0.57	0.06	0.61	0.22
F-*Wl*	0.55	0.36	0.48	0.51	0.49	0.64	0.41	0.44	0.53	0.58	0.44	0.32	0.54	0.44	0.38	0.49	0.60	0.17	0.22	0.14	0.23	0.56	0.41	0.09
M-*Wl*	0.45	0.37	0.33	0.34	0.31	0.53	0.34	0.39	0.38	0.33	0.53	0.53	0.45	0.45	0.41	0.54	0.52	0.38	0.51	0.32	0.46	0.44	0.52	0.18
F-*AvgS*	1.65	1.93	1.73	1.02	1.63	1.35	0.90	1.36	1.51	1.41	1.37	1.34	1.67	1.16	1.39	1.02	0.96	0.39	0.79	0.43	0.54	1.68	1.64	0.41
M-*AvgS*	1.83	1.34	1.32	1.19	1.00	1.53	1.61	0.99	1.22	1.14	1.13	0.94	1.30	1.45	1.35	1.33	1.40	1.60	0.55	1.13	1.43	0.91	1.09	1.15

The experimental results presented can be used to support analysis of relations between discretisation approaches, properties of induced decision rules, and individual attributes, enhancing understanding of existing dependencies.

5. Conclusions

The paper reports research on dependencies between discretisation methods, rule sets inferred from data, and individual characteristics obtained for attributes, built around rules. In the performed experiments the input datasets were discretised by selected approaches, and for all variants of these sets decision rules were induced in rough set processing. Properties of rules and rule sets, such as cardinalities, rule lengths and supports, were used as a base for definitions of several factors describing considered condition attributes. Statistics obtained for variables were analysed in relation to discretisation methods, which led to construction of the modified variants of learning sets, combining different discretisation approaches for different variables. From these new datasets decision rules were induced again, and the resulting characteristics of rules, sets, and attributes studied. The proposed methodology enabled deeper investigation of relationships existing between discretisation approaches and properties of inferred decision rules, visible through attributes and their characteristics, enhancing understanding of knowledge represented by rules.

Future research paths will include explorations of dependence of rule classifier performance on discretisation approaches and attribute characteristics. In particular combinations of various discretisation procedures within one dataset, and the influence of such processing on predictive powers of inducers will be studied.

Acknowledgements

The research works presented in the paper were performed at the University of Silesia in Katowice, and at the Silesian University of Technology, Gliwice, within the statutory project (RAU-6, 2021).

References

[1] J. Dougherty, R. Kohavi, M. Sahami, Supervised and unsupervised discretization of continuous features, in: Machine Learning: Proceedings of the 12th International Conference, Morgan Kaufmann, 1995, pp. 194–202.
[2] Y. Yang, G. I. Webb, X. Wu, Discretization Methods, in: O. Maimon, L. Rokach (Eds.), Data Mining and Knowledge Discovery Handbook, Springer US, Boston, MA, 2005, pp. 113–130.
[3] U. Stańczyk, B. Zielosko, G. Baron, Discretisation of conditions in decision rules induced for continuous data, PLOS ONE 15 (4) (2020) 1–33.
[4] S. García, J. Luengo, J. A. Sáez, V. López, F. Herrera, A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning, IEEE Transactions on Knowledge and Data Engineering 25 (4) (2013) 734–750.
[5] T. Amin, I. Chikalov, M. Moshkov, B. Zielosko, Dynamic programming approach for partial decision rule optimization, Fundamenta Informaticae 119 (3-4) (2012) 233–248.
[6] U. Stańczyk, Evaluating importance for numbers of bins in discretised learning and test sets, in: I. Czarnowski, J. R. Howlett, C. L. Jain (Eds.), Intelligent Decision Technologies 2017: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies (KES-IDT 2017) – Part II, Vol. 72 of Smart Innovation, Systems and Technologies, Springer International Publishing, 2018, pp. 159–169.
[7] Z. Pawlak, Rough sets and intelligent data analysis, Information Sciences 147 (2002) 1–12.
[8] M. Koppel, J. Schler, S. Argamon, Computational methods in authorship attribution, Journal of the American Society for Information Science and Technology 60 (1) (2009) 9–26.
[9] Y. Zhao, J. Zobel, Searching with style: Authorship attribution in classic literature, in: Proceedings of the Thirtieth Australasian Conference on Computer Science - Volume 62, ACSC '07, Australian Computer Society, Inc., Darlinghurst, Australia, 2007, pp. 59–68.
[10] L. Wróbel, M. Sikora, M. Michalak, Rule quality measures settings in classification, regression and survival rule induction — an empirical approach, Fundamenta Informaticae 149 (2016) 419–449.
[11] E. Stamatatos, A survey of modern authorship attribution methods, Journal of the Americal Society for Information Science and Technology 60 (3) (2009) 538–556.
[12] J. Rybicki, M. Eder, D. Hoover, Computational stylistics and text analysis, in: C. Crompton, R. Lane, R. Siemens (Eds.), Doing Digital Humanities: Practice, Training, Research, 1st Edition, Routledge, 2016, pp. 123–144.
[13] R. Dash, R. L. Paramguru, R. Dash, Comparative analysis of supervised and unsupervised discretization techniques, International Journal of Advances in Science and Technology 2 (3) (2011) 29–37.
[14] U. Fayyad, K. Irani, Multi-interval discretization of continuous valued attributes for classification learning, in: Proceedings of the 13th International Joint Conference on Artificial Intelligence, Vol. 2, Morgan Kaufmann Publishers, 1993, pp. 1022–1027.
[15] I. Kononenko, On biases in estimating multi-valued attributes, in: 14th International Joint Conference on Articial Intelligence, 1995, pp. 1034–1040.
[16] G. Baron, On sequential selection of attributes to be discretized for authorship attribution, in: 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), 2017, pp. 229–234.
[17] S. Kotsiantis, D. Kanellopoulos, Discretization techniques: A recent survey, International Transactions on Computer Science and Engineering 1 (32) (2006) 47–58.
[18] U. Stańczyk, On unsupervised and supervised discretisation in mining stylometric features, in: A. Gruca, T. Czachórski, S. Deorowicz, K. Har eżlak, A. Piotrowska (Eds.), Man-Machine Interactions 6. ICMMI 2019, Vol. 1061 of Advances in Intelligent Systems and Computing, Springer, Cham, 2020, pp. 156–166.
[19] J. W. Grzymala-Busse, Discretization based on entropy and multiple scanning, Entropy 15 (5) (2013) 1486–1502.
[20] U. Stańczyk, B. Zielosko, On approaches to discretisation of stylometric data and conflict resolution in decision making, in: I. J. Rudas, J. Csirik, C. Toro, J. Botzheim, R. J. Howlett, L. C. Jain (Eds.), Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 23rd International Conference KES-2019, Vol. 159 of Procedia Computer Science, Elsevier, 2019, pp. 1811–1820.
[21] R. Bello, Y. Gómez, Y. Caballero, A. Nowe, R. Falcón, Rough sets and evolutionary computation to solve the feature selection problem, in: A. Abraham, R. Falcón, R. Bello (Eds.), Rough Set Theory: A True Landmark in Data Analysis, Springer Berlin Heidelberg, 2009, pp. 235–260.
[22] F. Alsolami, T. Amin, M. Moshkov, B. Zielosko, K. Żabiński, Comparison of heuristics for optimization of association rules, Fundamenta Informaticae 166 (1) (2019) 1–14.
[23] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (1) (2007) 3–27.
[24] J. Stefanowski, D. Vanderpooten, Induction of decision rules in classification and discovery-oriented perspectives, International Journal of Intelligent Systems 16 (1) (2001) 13–27.
[25] A. Gudyś, M. Sikora, L. Wróbel, Rulekit: A comprehensive suite for rule-based learning, Knowledge-Based Systems 194 (2020) 105480.
[26] B. Zielosko, Application of dynamic programming approach to optimization of association rules relative to coverage and length, Fundamenta Informaticae 148 (1-2) (2016) 87–105.
[27] J. Bazan, M. Szczuka, The rough set exploration system, in: J. F. Peters, A. Skowron (Eds.), Transactions on Rough Sets III, Vol. 3400 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2005, pp. 37–56.