An Introduction of the Condition Class Space with Continuous Value Discretization and Rough Set Theory

Malcolm J. Beynon*
Cardiff Business School, Cardiff University, Colum Drive, Cardiff CF10 3EU, Wales, United Kingdom

The granularity of an information system has an incumbent effect on the efficacy of the analysis from many machine learning algorithms. An information system contains a universe of objects characterized and categorized by condition and decision attributes. To manage the concomitant granularity, a level of continuous value discretization (CVD) is often undertaken. In the case of the rough set theory (RST) methodology for object classification, the granularity contributes to the grouping of objects into condition classes with the same condition attribute values. This article exposits the effect of a level of CVD on the subsequent condition classes constructed, with the introduction of the condition class space—the domain within which the condition classes exist. This domain elucidates the association of the condition classes to the related decision outcomes—reflecting the inexactness incumbent when a level of CVD is undertaken. A series of measures is defined that quantify this association. Throughout this study and without loss of generality, the findings are made through the RST methodology. This further offers a novel exposition of the relationship between all the condition attributes and the RST-related reducts (subsets of condition attributes).

1. INTRODUCTION

Rough set theory (RST), introduced in Pawlak[1,2] is a nascent technique for data mining, in particular object classification. Central to its application is an information system that contains a universe of objects, each characterized and categorized by condition and decision attributes, respectively. With the use of an indiscernibility relation, condition and decision classes are constructed that group objects together with the same condition or decision attribute values, respectively. The outcome of an RST analysis is a set of "if . . . then . . ." decision rules, whose sets of conditions are the values that discern the concomitant condition classes. Reported advantages of the utilization of RST include its very clear interpretation for the user and its independence to any statistical assumptions.[3] Inspiration for

*e-mail: BeynonMJ@cardiff.ac.uk.

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. 21, 173–191 (2006)
 Published online in Wiley InterScience (www.interscience.wiley.com). • DOI 10.1002/int.20126

the utilization of RST is suggested by Dünstch and Gediga[4] (p. 594), through the motto “Let the data speak for themselves.”

Since its introduction, a number of issues relating to RST have been elucidated. One of these issues, incumbent in many machine learning algorithms, relates to the granularity of the information system considered. The granularity describes the level of grouping of objects into condition classes. Within RST, this granularity affects the number and specificity of the decision rules constructed. To illustrate, a high granularity may result in a large number of decision rules being constructed, possibly each describing only one or two objects. This large specificity can affect the interpretability (generality) exhibited in the conditions of the decision rules. The surrounding argument is described by the well-known science tenet—Occam’s Razor.[5] In summary, it relates to the possible preference for simpler models, which here may equate to the reduction in the number of conditions in the rules and actual rules in a constructed rule set.

To allow more opportunity for interpretability, a level of continuous value discretization (CVD) is often employed to reduce the granularity of the associated information system.[6,7] The CVD constructs intervals over the domain of a “continuous” condition attribute, for which values of objects in an interval are considered the same. It could be argued that the utilization of a level of CVD goes against the philosophy of the motto previously given,[4] because now an object’s original attribute values (data) are described by a number of general interval labels. In many cases the actual CVD is a subjective process undertaken by an expert, and as such may include a level of bias and/or overconfidence in the operation,[8] which would impose the “voice” of the expert.

To alleviate expert bias, automated (objective) methods of CVD have been developed that attempt to effectively group values over a known condition attribute domain. The general structure of these methods can be broken down into whether they are global or local, supervised or unsupervised, and static or dynamic.[9–11] Indeed, CVD methods based on RST have been constructed.[12,13] Recently, Beynon[14] introduced a number of stability measures to quantify the effectiveness of the CVD of continuous condition attributes, irrespective of the CVD method employed. These measures utilize the spread of the original values in each constructed interval. That is, although all the values in an interval are considered the same, certain of these values may be nearer to values in the neighboring intervals than those in its own interval. The study applied these measures in the selection of RST related reducts with most stability (see Section 2); importantly, the reduct selection criteria described was based *post* rule construction.

In this article, the general CVD stability measures are used to investigate the certainty of the classification of objects in the constructed condition classes. Moreover, for a condition class (of objects), a set of likelihood values are evaluated that describe its classification to each decision class (outcome). These sets of values elucidate the level of clustering of condition classes that are associated with the same decision outcome. The larger the likelihood values of condition classes to their correct (known) decision outcome the more stable their classification. When these likelihood values are considered as a vector, the notion of a condition class space is defined, for which a condition class is represented by a point in this space.

The findings are considered on the whole set and subsets of condition attributes, as such have a bearing on the elucidation of reducts in RST. Indeed, this study offers a novel approach to the exposition of condition classes associated with a reduct and their related condition classes from the whole set of condition attributes. The well-known Iris and wine data sets are utilized to exposit the methodology introduced here. As two "three-decision outcome" problems, the simplex plot representation of data is used to further exposit the innovative analysis presented in this article.

The structure of the rest of the article is as follows: In Section 2, the fundamentals of RST are presented as well as stability in relation to CVD. In Section 3, the notions of the classification stability and the condition class space are presented through the vectorization of the association of each condition class to the decision outcomes in this space. In Sections 4 and 5, the methodology presented is applied to the well-known Iris and wine data sets. In Section 6, conclusions are given as well as directions for future research.

2. FUNDAMENTALS OF ROUGH SET THEORY AND CONTINUOUS VALUE DISCRETIZATION

Central to the domain of an RST analysis is an information system or decision table. An information system is made up of a universe of objects (U), each characterized by a set of condition attributes (C) and categorized by a set of decision attributes (D). The nature (value) of an attribute value to an object is a descriptor value. From C and D certain equivalence classes (condition $E(C)$ and decision $E(D)$) are constructed through the utilization of an indiscernibility relation—where objects in a condition or decision class have the same series of relevant descriptor values. RST allows the association of the objects in U to a decision outcome $Y_i \in E(D)$ based on a subset of condition attributes $P \subseteq C$ to be described in terms of the pair of sets; lower approximation $\underline{PY_i}$ and upper approximation $\overline{PY_i}$, more formally defined by:

$$\underline{PY_i} = \{o_j \in U \,|\, o_j \in X_p \in E(P) \text{ and } X_p \subseteq Y_i\}$$

$$\overline{PY_i} = \{o_j \in U \,|\, o_j \in X_p \in E(P) \text{ and } X_p \cap Y_i \neq \varnothing\}$$

In words, the lower approximation is the union of all those condition classes ($X_p \in E(P)$) that are contained in Y_i and the upper approximation is the union of all those condition classes that have nonempty intersection with Y_i. From their definition the objects in a $\underline{PY_i}$ have a definite classification to the respective decision outcomes. A measure denoting the quality of classification is defined by $\gamma(P, D) = \sum_{i=1}^{|E(D)|} |\underline{PY_i}|/|U|$ and represents the proportion of objects from U that have a definite classification. This measure aids the identification of subsets of condition attributes P that have the same (or near same) level of classification as C, defined reducts.[2,15] It is from a reduct that the respective decision rules are constructed.

Throughout the rest of this article the case of when there are q condition attributes $c_1, c_2, \ldots, c_q$ (sets of continuous values) in an information system are considered, which have been intervalized by some predetermined CVD process. This assumption is without loss of generality, with the proposed findings applicable to the case of when an information system also contains nominal condition attributes. The labels for the intervals constructed become the associated descriptor values for the objects whose original values are in the specific intervals. In general, the jth interval of the nth condition attribute $c_n (1 \leq n \leq q)$ is defined as I_j^n; also $\mathrm{rgt}(\mathrm{I}_j^n)$ and $\mathrm{lft}(\mathrm{I}_j^n)$ denote its right and left boundary points. Given c_n has been discretized into k_n intervals, if we consider I_r^n, then from Beynon[14] the proportion $S_{n,j,r}$ of the estimated distribution constructed from the original data in the jth interval actually in the rth interval is given by

$$S_{n,j,r} = \int_{\mathrm{lft}(\mathrm{I}_r^n)}^{\mathrm{rgt}(\mathrm{I}_r^n)} \frac{1}{\sqrt{2m_{n,j}\pi}} \sum_{i=1}^{m_{n,j}} \frac{1}{\mathrm{rgt}(\mathrm{I}_j^n) - \mathrm{lft}(\mathrm{I}_j^n)} \times \exp\left[-\frac{m_{n,j}}{2}\left(\frac{x - x_i}{\mathrm{rgt}(\mathrm{I}_j^n) - \mathrm{lft}(\mathrm{I}_j^n)}\right)^2\right] dx$$

where $m_{n,j}$ is the number of objects whose original values lie in the I_j^n interval. The sum of the $S_{n,j,r}$ values should equal one ($\sum_{r=1}^{k_n} S_{n,j,r} = 1$), but because each estimated distribution exists over the domain $(-\infty, \infty)$, not all of this domain may be attainable by the condition attribute in question. Consideration has to be given to the extreme values $\mathrm{lft}(\mathrm{I}_1^n)$ and $\mathrm{rgt}(\mathrm{I}_{k_n}^n)$. Where defined extreme values exist, there is no problem (e.g., percentage with 0 and 100), for undefined extremes; the $\mathrm{lft}(\mathrm{I}_1^n)$ and $\mathrm{rgt}(\mathrm{I}_{k_n}^n)$ are here defined as the minimum and maximum assigned values in I_1^n and $\mathrm{I}_{k_n}^n$, respectively. Importantly, in the case of finite extreme values for a condition attribute, the $S_{n,j,r}$ values should be normalized.

A value $S_{n,j,r}$ $(1 \leq j, r \leq k_n)$ is the likelihood of an object's value from the nth condition attribute contained in the jth interval should actually be contained in the rth interval. In terms of descriptor values, defined as $\delta_{n,j}, j = 1, \ldots, k_n$, on c_n, for an object described by $\delta_{n,j}$, $S_{n,j,r}$ is the likelihood that it should be described by $\delta_{n,r}$.

3. CLASSIFICATION STABILITY AND CONDITION CLASS SPACE

This section considers the effect of a level of CVD undertaken on a subset of condition attributes, resulting in a concomitant number of condition classes. Moreover, a subset of condition attributes, defined as $s \subseteq C$, is made up of the condition attributes $c_h^s (h = 1, \ldots, |s|)$. With a level of CVD inherent, based on s, the universe of objects is partitioned into a number of condition classes $X_q^s (q = 1, \ldots, |E(s)|)$. Each condition class X_q^s is defined by a distinct series (list) of condition attribute descriptor values, defined as $[\delta_{q,1}^s, \delta_{q,2}^s, \ldots, \delta_{q,|s|}^s]$. To reiterate, each term $\delta_{q,h}^s (h = 1, \ldots, |s|)$ denotes the descriptor value of the hth condition attribute for an object in the qth condition class associated with the subset of condition attributes s.

For an object $o_j \in U$, its set of descriptor values would associate it with a particular condition class, say X_q^s, denoted by $o_j \in X_q^s = [\delta_{q,1}^s, \delta_{q,s}^s, \ldots, \delta_{q,|s|}^s]$. In the presence of CVD, for each condition attribute there exists levels of likelihood in which descriptor value an object value could be associated with (see Section 2).[14] Furthermore, given the actual condition class an object is contained in ($o_j \in X_q^s$), the individual component likelihood values that describe it and all objects in the condition class X_q^s as possibly being in another condition class, say X_p^s, is given by $[S_{c_1^s,\delta_{q,1}^s,\delta_{p,1}^s}, S_{c_2^s,\delta_{q,2}^s,\delta_{p,2}^s}, \ldots, S_{c_{|s|}^s,\delta_{q,|s|}^s,\delta_{p,|s|}^s}]$. A $S_{c_h^s,\delta_{q,h}^s,\delta_{p,h}^s}$ value is the likelihood that an object's original value from c_h^s, denoted by the descriptor value $\delta_{q,h}^s$ should be described by the descriptor value $\delta_{p,h}^s$.

These are component likelihood values, which need to be aggregated together to quantify a condition class's (X_q^s) possible transference of classification (of objects) to that associated with another condition class, say X_p^s. Because the membership of an object to a condition class is a conjunction based on the associated descriptor values, a geometric mean (aggregation) value is utilized. It follows that the likelihood that an object contained in the condition class X_q^s should be in X_p^s is defined as $\mathrm{CT}_{q,p}^s$ (CT: condition transference) and given by

$$\mathrm{CT}_{q,p}^s = \left(\prod_{h=1}^{|s|} S_{c_h^s,\delta_{q,h}^s,\delta_{p,h}^s} \right)^{1/|s|}, \quad 1 \le q, p \le |E(s)|$$

The $\mathrm{CT}_{q,p}^s$ $1 \le q, p \le |E(s)|$ can also be considered the containment likelihood values of objects in all the condition classes to all other condition classes, associated with a subset of condition attributes s. The importance of these "containment" values is apparent at the decision class level. That is, a subset of condition attributes defines a series of condition classes including objects classified to a number of decision outcomes. Again, without loss of generality, but following the deterministic nature of RST, here only condition classes that each contain objects to single decision outcomes are further considered.

For the condition classes considered, they have $\mathrm{CT}_{q,p}^s$ values associating them with the different decision outcomes through their direct association with the other condition classes. Using this relationship, a measure defined as $\mathrm{DT}_{q,i}^s$ (DT: decision transference) is presented that quantifies the association of a condition class X_q^s to a decision outcome, say $Y_i \in E(D)$, that includes evidence from all condition classes, and is given by

$$\mathrm{DT}_{q,i}^s = \frac{\sum\limits_{X_p^s \subseteq Y_i} \mathrm{CT}_{q,p}^s}{\sum\limits_{j=1}^{|E(D)|} \left(\sum\limits_{X_p^s \subseteq Y_j} \mathrm{CT}_{q,p}^s \right)} = \frac{\sum\limits_{X_p^s \subseteq Y_i} \left(\prod\limits_{h=1}^{|s|} S_{c_h^s,\delta_{q,h}^s,\delta_{p,h}^s} \right)^{1/|s|}}{\sum\limits_{j=1}^{|E(D)|} \left(\sum\limits_{X_p^s \subseteq Y_j} \left(\prod\limits_{h=1}^{|s|} S_{c_h^s,\delta_{q,h}^s,\delta_{p,h}^s} \right)^{1/|s|} \right)}$$

for $1 \le q \le |E(s)|$ and $1 \le i \le |E(D)|$. A series of these values can be constructed, $\mathrm{DT}_{q,1}^s, \mathrm{DT}_{q,2}^s, \ldots, \mathrm{DT}_{q,|E(D)|}^s$, where $\mathrm{DT}_{q,i}^s$ is the likelihood that a condition class X_q^s found from the subset of condition attributes $s \subseteq C$ should be associated with a decision outcome $Y_i \in E(D)$. This expression allows each condition class to provide equal evidence, subject to their $\mathrm{CT}_{q,p}^s$ values, to each other's association with all the possible decision outcomes. However, each condition class contains different numbers of objects, which indicates the importance (relevance) of the individual condition classes. To include a measure of the relevance of each condition class, a series of weighted $\mathrm{DT}^s_{q,i}$ values, defined as $w\mathrm{DT}^s_{q,i}$, are constructed and given by

$$w\mathrm{DT}^s_{q,i} = \frac{\sum\limits_{X^s_p \subseteq Y_i} (\ln|X^s_p|\mathrm{CT}^s_{q,p})}{\sum\limits_{j=1}^{|E(D)|}\left(\sum\limits_{X^s_p \subseteq Y_j} (\ln|X^s_p|\mathrm{CT}^s_{q,p})\right)}$$

$$= \frac{\sum\limits_{X^s_p \subseteq Y_i}\left(\ln|X^s_p|\left(\prod\limits_{h=1}^{|s|} S_{c^s_h,\delta^s_{q,h},\delta^s_{p,h}}\right)^{1/|s|}\right)}{\sum\limits_{j=1}^{|E(D)|}\left(\sum\limits_{X^s_p \subseteq Y_j}\left(\ln|X^s_p|\left(\prod\limits_{h=1}^{|s|} S_{c^s_h,\delta^s_{q,h},\delta^s_{p,h}}\right)^{1/|s|}\right)\right)}$$

The $\ln|X^s_p|$ term is a level of grouping of the objects. In the limit, a condition class containing one object (possible outlier) would mean $\ln|X^s_p| = 0$; hence the condition class contributes nothing in support to the classification of the condition class in question. As the size of a condition class increases, so $\ln|X^s_q|$ also increases, but not linearly. In summary, the expressions presented in this section are a consequence of the need to have undertaken a level of CVD on the condition attributes in an information system. The $\mathrm{DT}^s_{q,i}$ and $w\mathrm{DT}^s_{q,i}$ $1 \le q \le |E(s)|$, $1 \le i \le |E(D)|$ values are constructed from the evidence of all the considered condition classes on the association of a condition class to each decision outcome. This possibility of supporting evidence from other condition classes is a consequence of the inexactness of interval association of object values in the intervals constructed from the CVD undertaken.

For a single condition class X^s_q, the list $[\mathrm{DT}^s_{q,1}, \mathrm{DT}^s_{q,2}, \ldots, \mathrm{DT}^s_{q,|E(D)|}]$ can be considered a vector, defined as $\overline{\mathrm{DT}^s_q}$ (similar for $\overline{w\mathrm{DT}^s_q}$), representing its association with all the decision outcomes. The respective series of $\overline{\mathrm{DT}^s_q}$ (or $\overline{w\mathrm{DT}^s_q}$) $q = 1, \ldots, |E(s)|$ vectors place the condition classes in relation to each other, in terms of their classification. Moreover, they place each condition class in the respective *condition class space* for the problem in question. For an information system with $|E(D)|$ decision outcomes, the optimum vectors $[1,0,0,\ldots,0,0], [0,1,0,\ldots,0,0], \ldots, [0,0,0,\ldots,0,1]$ would define where a condition class has definite (certain) association to a single decision outcome.

This vectorization allows one more exposition of the effectiveness of the CVD undertaken. That is, it allows the quantification of the level of overall certainty in the correct classification of the objects in the information system (post CVD). This quantification is simply the aggregations of the distances of each condition class vector to their optimum representation—certainty in classification. Hence a measure of the influence of CVD is the overall distances of the $\overline{\mathrm{DT}^s_q}$ or $\overline{w\mathrm{DT}^s_q}$, $q = 1, \ldots, |E(s)|$ vectors for each condition class X^s_q from their respective optimum vector. Moreover, a mean Euclidean distance could be used in each case, defined as ξ^s and ξ^s_w, which is the mean of the respective distances.

Table I. Intervalization of the condition attributes in the Iris data set.[16]

	Very small (1)	Small (2)	Large(3)	Very Large (4)
c_1	$x < 50$ (22)	$50 \le x < 60$ (61)	$60 \le x < 70$ (54)	$70 \le x$ (13)
c_2	$x < 24$ (8)	$24 \le x < 31$ (75)	$31 \le x < 38$ (55)	$38 \le x$ (12)
c_3	$x < 30$ (50)	$30 \le x < 40$ (11)	$40 \le x < 55$ (61)	$55 \le x$ (28)
c_4	$x < 10$ (50)	$10 \le x < 14$ (28)	$14 \le x < 21$ (49)	$21 \le x$ (23)

4. APPLICATION OF THE CONDITION CLASS STABILITY ANALYSIS WITH THE IRIS DATA SET

In this section, the findings presented in Section 3 are applied to the well-known Iris data set. This data set is made up of three classes of 50 plants each, where a decision class (outcome) refers to a particular type of Iris plant (i.e., *Iris Setosa*, *Iris Versicolour*, and *Iris Virginica*). Four continuous characteristics (condition attributes) are used to describe each plant, namely sepal length (c_1), sepal width (c_2), petal length (c_3), and petal width (c_4). For brevity we consider the paper[16] that includes a level of CVD on the Iris data set to reduce the overall granularity of the associated information system (see Table I).

In Table I, the interval boundary values defining the CVD of each condition attribute are reported.[3,16] Also presented (in parentheses) are the number of original values in each defined interval. Following Ref. 14, a series of estimated distributions can be constructed for each set of objects (plants) in the intervals. This uses the method of Parzen windows,[17] with its functional form included in the expression for $S_{n,j,r}$ (in Section 2), the results are reported in Figure 1.

In Figure 1, the graphs of the constructed estimated distributions are presented; also included are vertical dashed lines defining the boundaries between intervals (from Table I). An initial inspection of the graphs show a level of overlap in the estimated distributions; this is due to their domains each being over (0, ∞) but highlights possible indecisiveness in the boundaries constructed. To illustrate the caution that should be taken when deciding on the interval boundary values, the intervals "3" and "4" of the c_2 attribute are further considered (see Figure 2).

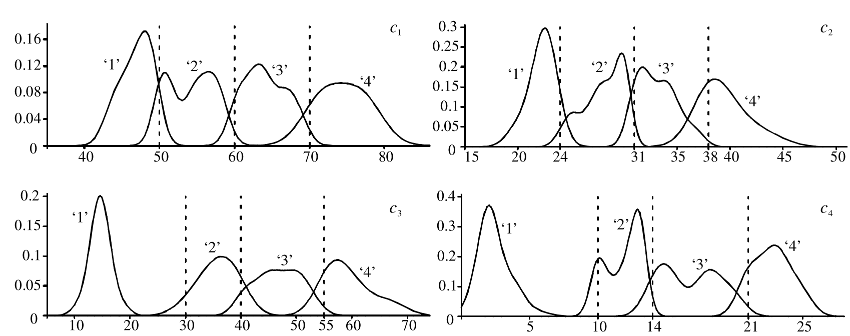

Figure 1. Estimated distributions of attribute values in each constructed interval.

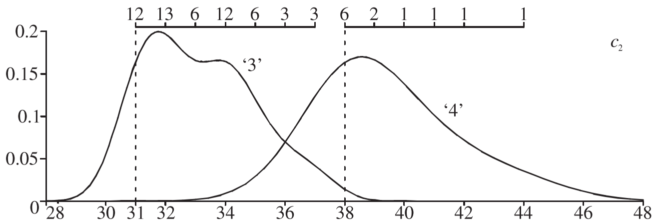

Figure 2. Detailed information on the "3" and "4" intervals of the c_2 condition attribute.

In Figure 2, the estimated distributions associated with the "3" and "4" intervals are presented with the positions and frequencies of the original values also shown (on the top axis). To illustrate, 12 plants have the value 31 mm for c_2 (sepal width). These values show that the boundary point between these intervals is at the lowest value in the "4" interval and away from the larger values in the "3" interval. This boundary value was chosen by an expert and its effect on the "4" interval is noticeable. That is, its estimated distribution spreads considerably into the domain of the "3" interval, a consequence of a majority of the values in the "4" interval being close to the (left) boundary value. This then offers a level of likelihood that values in the "4" interval should be in the "3" interval. This indecisiveness can be quantified, with the $S_{n,j,r}$ values associated with each condition attribute intervalized (see Table II).

In Table II, the values (in bold) on the leading diagonals represent the stability ($S_{n,j,j}$ values) of the intervals in question. That is, the likelihood that an object's original value is in the correct interval (described by the correct descriptor value). The lowest of these is associated with I_4^2, the "4" interval of the c_2 attribute ($S_{2,4,4} = 0.6883$). Explanation for this low value is given, in part, with the discussion accompanying Figure 2 presented earlier. The other values $S_{n,j,r}$ ($j \neq r$) are the likelihood that an object's original value should be in another (neighboring) interval. Understandably these other likelihood values are largest with the immediate neighbor intervals of the $S_{n,j,j}$ values. All the $S_{n,j,r}$ values are then used to define the levels of classification of each condition class.

Table II. $S_{n,j,r}$ values for the different intervals describing condition attributes.

		Intervals						Intervals			
$S_{n,j,r}$ value		"1"	"2"	"3"	"4"	$S_{n,j,r}$ value		"1"	"2"	"3"	"4"
c_1	"1"	**0.9082**	0.0918	0.0000	0.0000	c_3	"1"	**1.0000**	0.0000	0.0000	0.0000
	"2"	0.1182	**0.8630**	0.0188	0.0000		"2"	0.0872	**0.7552**	0.1576	0.0000
	"3"	0.0000	0.0890	**0.8878**	0.0232		"3"	0.0000	0.0694	**0.9111**	0.0195
	"4"	0.0000	0.0000	0.1282	**0.8718**		"4"	0.0000	0.0000	0.1751	**0.8249**
c_2	"1"	**0.8503**	0.1497	0.0000	0.0000	c_4	"1"	**1.0000**	0.0000	0.0000	0.0000
	"2"	0.0320	**0.9297**	0.0384	0.0000		"2"	0.1357	**0.8204**	0.0439	0.0000
	"3"	0.0000	0.1453	**0.8458**	0.0089		"3"	0.0001	0.1224	**0.8555**	0.0221
	"4"	0.0000	0.0000	0.3117	**0.6883**		"4"	0.0000	0.0000	0.1484	**0.8516**

Using these constructed intervals with all condition attributes C $(=\{c_1, c_2, c_3, c_4\})$ there exist 34 condition classes. Of these, 30 include objects categorized to the same decision outcomes within an individual condition class. The associated quality of classification is $\gamma(C, D) = 0.7733$, indicating 116 out of the 150 plants are assigned a definite classification. Concentrating on these 30 condition classes, their respective $\mathrm{CT}^s_{q,p}$ $1 \le q, p \le |C|$ values are next found (a total of 900 individual values). Due to the number of these values, they are not presented. Moving on from the containment likelihood values, the association of condition classes to each of the decision outcomes is investigated. Table III reports the $\overline{\mathrm{DT}_q^C}$ and $\overline{w\mathrm{DT}_q^C}$ $(q = 1, \ldots, 30)$ vectors for the 30 condition classes considered.

In Table III, for each of the condition classes the descriptor values that define it and the number of objects it contains are also presented. With three decision outcomes, each condition class is described by a vector $\overline{\mathrm{DT}_q^C} = [\mathrm{DT}^C_{q,1}, \mathrm{DT}^C_{q,2}, \mathrm{DT}^C_{q,3}]$ (similarly for $\overline{w\mathrm{DT}_q^C}$). Inspection of these vectors shows the specific $\mathrm{DT}^C_{q,i}$ value of the actual (known) decision outcome each condition class is

Table III. Stability details on condition classes X_q^C, $q = 1, \ldots, 30$.

q	$E(D)$	$X_q^C : \lvert X_q^C \rvert$	$[\mathrm{DT}^C_{q,1}, \mathrm{DT}^C_{q,2}, \mathrm{DT}^C_{q,3}]$	$[w\mathrm{DT}^C_{q,1}, w\mathrm{DT}^C_{q,2}, w\mathrm{DT}^C_{q,3}]$
1	1	[1, 1, 1, 1]: 1	**0.9980**, 0.0020, 0.0000	**0.9976**, 0.0024, 0.0000
2	1	[1, 2, 1, 1]: 6	**0.9983**, 0.0017, 0.0000	**0.9991**, 0.0009, 0.0000
3	2	[1, 2, 2, 2]: 1	0.2487, **0.6707**, 0.0000	0.3907, **0.6093**, 0.0000
4	3	[1, 2, 3, 3]: 1	0.0005, 0.5704, **0.4291**	0.0015, 0.9982, **0.0004**
5	1	[1, 3, 1, 1]: 13	**0.9991**, 0.0009, 0.0001	**0.9997**, 0.0003, 0.0000
6	2	[2, 1, 2, 2]: 2	0.1331, **0.8390**, 0.0279	0.0748, **0.9246**, 0.0006
7	2	[2, 1, 3, 2]: 1	0.0027, **0.9347**, 0.0626	0.0014, **0.9791**, 0.0195
8	1	[2, 2, 1, 1]: 1	**0.9980**, 0.0020, 0.0000	**0.9981**, 0.0019, 0.0000
9	2	[2, 2, 2, 2]: 7	0.1866, **0.7745**, 0.0389	0.2125, **0.7869**, 0.0006
10	2	[2, 2, 2, 3]: 1	0.0175, **0.7865**, 0.1960	0.0322, **0.9408**, 0.0270
11	2	[2, 2, 3, 2]: 10	0.0041, **0.8954**, 0.1005	0.0045, **0.9729**, 0.0226
12	3	[2, 2, 3, 4]: 1	0.0000, 0.2150, **0.7850**	0.0000, 0.0000, **1.0000**
13	1	[2, 3, 1, 1]: 19	**0.9991**, 0.0009, 0.0000	**0.9995**, 0.0005, 0.0000
14	2	[2, 3, 3, 3]: 1	0.0004, **0.6412**, 0.3584	0.0014, **0.7520**, 0.2466
15	1	[2, 4, 1, 1]: 10	**0.9999**, 0.0001, 0.0000	**0.9999**, 0.0001, 0.0000
16	2	[3, 1, 3, 2]: 2	0.0007, **0.9485**, 0.0508	0.0000, **0.9315**, 0.0685
17	2	[3, 2, 3, 2]: 5	0.0015, **0.8881**, 0.1104	0.0011, **0.9143**, 0.0846
18	3	[3, 2, 3, 4]: 1	0.0000, 0.1253, **0.8747**	0.0000, 0.0000, **1.0000**
19	3	[3, 2, 4, 3]: 4	0.0000, 0.2665, **0.7335**	0.0000, 0.2926, **0.7074**
20	3	[3, 2, 4, 4]: 4	0.0000, 0.0393, **0.9607**	0.0000, 0.0000, **1.0000**
21	3	[3, 3, 3, 4]: 4	0.0000, 0.1741, **0.8259**	0.0000, 0.0000, **1.0000**
22	3	[3, 3, 4, 3]: 1	0.0000, 0.2142, **0.7858**	0.0000, 0.2121, **0.7879**
23	3	[3, 3, 4, 4]: 7	0.0000, 0.0733, **0.9267**	0.0000, 0.0000, **1.0000**
24	2	[4, 2, 4, 3]: 4	0.0000, **0.1677**, 0.8323	0.0000, **0.1336**, 0.8664
25	3	[4, 2, 4, 4]: 4	0.0000, 0.0438, **0.9562**	0.0000, 0.0000, **1.0000**
26	2	[4, 3, 3, 3]: 1	0.0000, **0.3970**, 0.6029	0.0000, **0.2571**, 0.7429
27	3	[4, 3, 4, 3]: 1	0.0000, 0.1657, **0.8343**	0.0000, 0.0971, **0.9029**
28	3	[4, 3, 4, 4]: 1	0.0000, 0.0753, **0.9247**	0.0000, 0.0000, **1.0000**
29	3	[4, 4, 4, 3]: 1	0.0000, 0.1440, **0.8560**	0.0000, 0.0369, **0.9631**
30	3	[4, 4, 4, 4]: 1	0.0000, 0.0788, **0.9212**	0.0000, 0.0000, **1.0000**

associated with is given in bold and the largest value in each vector is underlined. For an illustration of the construction of a vector, see later in this section. A brief inspection of the vectors shows three $\overline{\mathrm{DT}}_q^C$ (q = 4, 24, and 26) and three $\overline{w\mathrm{DT}}_q^C$ (q = 4, 24, and 26) that have different values in bold and underlined. These incongruencies identify where the majority association to a decision outcome based on "neighbor" condition classes (because of CVD) is different from its known classification. This highlights the possible detrimental impact of the CVD undertaken.

When the condition classes are considered through these vectors, their relationship to each other requires an understanding of the domain they can exist in. The domain in this case is defined as the condition class space (see Section 3). For the Iris data set, each vector sums to one (e.g., $\sum_{i=1}^{3} \mathrm{DT}_{q,i}^C = 1$); with three decision outcomes the condition class space (using $\overline{\mathrm{DT}}_q^C$ or $\overline{w\mathrm{DT}}_q^C$) can be represented by a simplex plot (see Figure 3).

In Figure 3a, the simplex plot includes simplex coordinates (denoted by circles) that each represent an ordered vector $\overline{\mathrm{DT}}_q^C = [\mathrm{DT}_{q,1}^C, \mathrm{DT}_{q,2}^C, \mathrm{DT}_{q,3}^C]$, $q = 1, \ldots, 30$ (similar for $\overline{w\mathrm{DT}}_q^C$; see Figure 3b). The dashed lines present the boundaries that partition where in a simplex plot a single value in a $\overline{\mathrm{DT}}_q^C$ vector is the largest, the "1," "2," and "3" labels at the vertices denoting the (ith) index of the (dominant) decision outcomes Y_1, Y_2, and Y_3, respectively. Hence, for correct classification, a condition class associated with the decision outcome $Y_i \in E(D)$ should be in the region, nearest to the ith labeled vertex.

The presented simplex plots elucidate a number of features associated with the $\overline{\mathrm{DT}}_q^C$ and $\overline{w\mathrm{DT}}_q^C$ vectors. First, the nearer a simplex coordinate is to a vertex the increased dominance of a single $\mathrm{DT}_{q,i}^C$ or $w\mathrm{DT}_{q,i}^C$ value is apparent, the more association of a condition class to a single decision outcome. Second, a simplex coordinate along an edge indicates an association to only two decision classes (from the three possible). In this case the lack of simplex coordinates inside the simplex plot shows a limited presence of association of condition classes to all three decision outcomes. This identifies in the condition classes considered that their clustering within the condition class space is mostly in groups that are associated with single or pairs of decision outcomes. To quantify the effect of the CVD

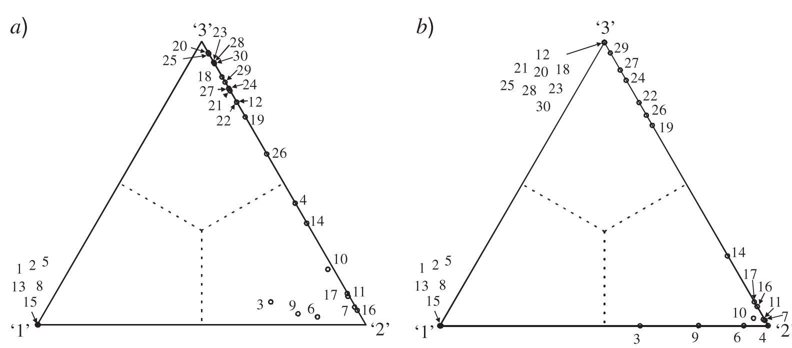

Figure 3. Simplex plot representation of the classification stability of each condition class.

and the condition class's classification to their correct decision outcomes, the ξ^C and ξ_w^C distance measures are considered. As a consequence of the CVD undertaken on all four condition attributes in the Iris data set, the resultant distance values are $\xi^C = 0.2128$ and $\xi_w^C = 0.1743$.

Until now the introduction of the condition class space has included the vectorization of the individual condition classes from C and their association to each decision outcome. We next consider the vectorization of condition class's classification associated with proper subsets of condition attributes, in particular reducts (in RST). In the investigation of reducts in this case, no proper subset ($P \subset C$) offers the same level of quality of classification as C. Here consideration is given to proper subsets of condition attributes that offer a similar level of quality of classification.[15] Investigation showed $\{c_1, c_3, c_4\}$ is a "near" reduct, defined as r_1, with $\gamma(\{c_1, c_3, c_4\}, D) = 0.7667$ (one less plant given a definite classification than with C). For this reduct it has 17 condition classes, of which 15 include objects categorized to the same decision outcome within their individual condition classes (see Table IV).

In Table IV, each of the 15 condition classes considered is defined based on its set of descriptor values and also the number of objects contained therein (ranges from 1 to 30). The number of objects in the condition classes sum to 115, which confirms the quality of classification value $\gamma(\{c_1, c_3, c_4\}, D) = 0.7667$ for the reduct $r_1 = \{c_1, c_3, c_4\}$. The $\mathrm{DT}_q^{r_1}$ and $w\mathrm{DT}_q^{r_1}$ vectors are also presented for each condition class. This reduct only has the condition attribute c_2 missing from the whole set of condition attributes. It follows that the condition classes X_q^C $q = 1, \ldots, 30$ may be individually associated with one of the condition classes $X_q^{r_1}$ $q = 1, \ldots, 15$, by the removal of the c_2 condition attribute. Hence, also shown in Table IV (last column) are the condition class indexes of X_q^C, which are associated with the respective $X_q^{r_1}$ condition classes. Inspection of these values shows the condition class X_{14}^C is not

Table IV. Stability details on condition classes $X_q^{r_1}$, $q = 1, \ldots, 15$.

q	$E(D)$	$X_q^{r_1} : \lvert X_q^{r_1} \rvert$	$[\mathrm{DT}_{q,1}^{r_1}, \mathrm{DT}_{q,2}^{r_1}, \mathrm{DT}_{q,3}^{r_1}]$	$[w\mathrm{DT}_{q,1}^{r_1}, w\mathrm{DT}_{q,2}^{r_1}, w\mathrm{DT}_{q,3}^{r_1}]$	X_q^C
1	1	[1, 1, 1]: 20	**0.9996**, 0.0004, 0.0000	**0.9999**, 0.0001, 0.0000	1, 2, 5
2	2	[1, 2, 2]: 1	0.1547, **0.7571**, 0.0882	0.4206, **0.5794**, 0.0000	3
3	3	[1, 3, 3]: 1	0.0000, 0.4024, **0.5976**	0.0002, 0.9998, **0.0000**	4
4	1	[2, 1, 1]: 30	**0.9996**, 0.0004, 0.0000	**0.9998**, 0.0002, 0.0000	8, 13, 15
5	2	[2, 2, 2]: 9	0.1277, **0.8359**, 0.0364	0.2511, **0.7488**, 0.0001	6, 9
6	2	[2, 2, 3]: 1	0.0050, **0.8008**, 0.1942	0.0201, **0.9430**, 0.0369	10
7	2	[2, 3, 2]: 11	0.0007, **0.9024**, 0.0969	0.0014, **0.9866**, 0.0120	7, 11
8	3	[2, 3, 4]: 1	0.0000, 0.1230, **0.8770**	0.0000, 0.0001, **0.9999**	12
9	2	[3, 3, 2]: 7	0.0003, **0.9313**, 0.0684	0.0005, **0.9393**, 0.0603	16, 17
10	3	[3, 3, 4]: 5	0.0000, 0.1198, **0.8802**	0.0000, 0.0000, **1.0000**	18, 21
11	3	[3, 4, 3]: 5	0.0000, 0.2467, **0.7533**	0.0000, 0.2250, **0.7750**	19, 22
12	3	[3, 4, 4]: 11	0.0000, 0.0331, **0.9669**	0.0000, 0.0000, **1.0000**	20, 23
13	2	[4, 3, 3]: 1	0.0000, **0.6456**, 0.3544	0.0000, **0.3131**, 0.6869	26
14	3	[4, 4, 3]: 6	0.0000, 0.2696, **0.7304**	0.0000, 0.0824, **0.9176**	24, 27, 29
15	3	[4, 4, 4]: 6	0.0000, 0.1097, **0.8903**	0.0000, 0.0000, **1.0000**	25, 28, 30

included in an $X_q^{r_1}$ condition class, the reason being that it is now included with "still" not considered condition classes of C (of which there are four).

For a single condition class, namely $X_{13}^{r_1}$, different $w\mathrm{DT}_{13,i}^{r_1}$ values within the $w\mathrm{DT}_{13}^{r_1}$ vector are in bold and underlined in Table IV. In the $w\mathrm{DT}_{13}^{r_1}$ vector, the most dominant decision outcome ($d_1 = 3$, underlined) it is associated with is different from its actual decision outcome ($d_1 = 2$, in bold). This condition class is made up of a single object, thus highlighting the possible outlier nature of it. With a manageable number of condition classes considered, the construction of the $w\mathrm{DT}_{13}^{r_1}$ vector $[w\mathrm{DT}_{13,1}^{r_1}, w\mathrm{DT}_{13,1}^{r_1}, w\mathrm{DT}_{13,3}^{r_1}]$ is next presented, with the numerator parts of the individual $w\mathrm{DT}_{13,i}^{r_1}$, $i = 1, 2$, and 3, expressions only given (note some 0.0000 values are small values):

$$\sum_{X_p^{r_1} \subseteq Y_1} \left(\ln|X_p^{r_1}| \left(\prod_{h=1}^{|r_1|} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{p,h}^{r_1}} \right)^{1/|r_1|} \right)$$

$$= \ln|X_1^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{1,h}^{r_1}} \right)^{1/3} + \ln|X_4^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{4,h}^{r_1}} \right)^{1/3}$$

$$= \ln|X_1^{r_1}|(S_{1,4,1} \times S_{3,3,1} \times S_{4,3,1})^{1/3} + \ln|X_4^{r_1}|(S_{1,4,2} \times S_{3,3,1} \times S_{4,3,1})^{1/3}$$

$$= \ln(20)(0.0000 \times 0.0000 \times 0.0001)^{1/3}$$

$$+ \ln(30)(0.0890 \times 0.0000 \times 0.0001)^{1/3}$$

$$= 2.9957 \times 0.0000 + 3.4012 \times 0.0000$$

$$= 0.0000 + 0.0000$$

$$= 0.0000$$

$$\sum_{X_p^{r_1} \subseteq Y_2} \left(\ln|X_p^{r_1}| \left(\prod_{h=1}^{|r_1|} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{p,h}^{r_1}} \right)^{1/|r_1|} \right)$$

$$= \ln|X_2^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{2,h}^{r_1}} \right)^{1/3} + \ln|X_5^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{7,h}^{r_1}} \right)^{1/3}$$

$$+ \ln|X_6^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{6,h}^{r_1}} \right)^{1/3} + \ln|X_7^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{7,h}^{r_1}} \right)^{1/3}$$

$$+ \ln|X_9^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{9,h}^{r_1}} \right)^{1/3} + \ln|X_{13}^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{13,h}^{r_1}} \right)^{1/3}$$

$$= \ln|X_2^{r_1}|(S_{1,4,1} \times S_{3,3,2} \times S_{4,3,2})^{1/3} + \ln|X_5^{r_1}|(S_{1,4,2} \times S_{3,3,2} \times S_{4,3,2})^{1/3}$$

$$+ \ln|X_6^{r_1}|(S_{1,4,2} \times S_{3,3,2} \times S_{4,3,3})^{1/3} + \ln|X_7^{r_1}|(S_{1,4,2} \times S_{3,3,3} \times S_{4,3,2})^{1/3}$$

$$+ \ln|X_9^{r_1}|(S_{1,4,3} \times S_{3,3,3} \times S_{4,3,2})^{1/3} + \ln|X_{13}^{r_1}|(S_{1,4,4} \times S_{3,3,3} \times S_{4,3,3})^{1/3}$$

$$
\begin{aligned}
&= \ln(1)(0.0000 \times 0.0694 \times 0.1224)^{1/3} \\
&\quad + \ln(9)(0.0000 \times 0.0694 \times 0.1224)^{1/3} \\
&\quad + \ln(1)(0.0000 \times 0.0694 \times 0.8555)^{1/3} \\
&\quad + \ln(11)(0.0000 \times 0.9111 \times 0.1224)^{1/3} \\
&\quad + \ln(7)(0.1282 \times 0.9111 \times 0.1224)^{1/3} \\
&\quad + \ln(1)(0.8718 \times 0.9111 \times 0.8555)^{1/3} \\
&= 0.0000 \times 0.0000 + 2.1972 \times 0.0029 + 0.0000 \times 0.0056 \\
&\quad + 2.3979 \times 0.0069 + 1.9459 \times 0.2427 + 0.0000 \times 0.8792 \\
&= 0.0000 + 0.0064 + 0.0000 + 0.0166 + 0.4723 + 0.0000 \\
&= 0.4953
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{X_p^{r_1} \subseteq Y_3} \left(\ln|X_p^{r_1}| \left(\prod_{h=1}^{|r_1|} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{p,h}^{r_1}} \right)^{1/|r_1|} \right) \\
&= \ln|X_3^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{3,h}^{r_1}} \right)^{1/3} + \ln|X_8^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{8,h}^{r_1}} \right)^{1/3} \\
&\quad + \ln|X_{10}^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{10,h}^{r_1}} \right)^{1/3} + \ln|X_{11}^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{11,h}^{r_1}} \right)^{1/3} \\
&\quad + \ln|X_{12}^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{12,h}^{r_1}} \right)^{1/3} + \ln|X_{14}^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{14,h}^{r_1}} \right)^{1/3} \\
&\quad + \ln|X_{15}^{r_1}| \left(\prod_{h=1}^{3} S_{c_h^{r_1}, \delta_{13,h}^{r_1}, \delta_{15,h}^{r_1}} \right)^{1/3} \\
&= \ln|X_3^{r_1}|(S_{1,4,1} \times S_{3,3,3} \times S_{4,3,3})^{1/3} + \ln|X_8^{r_1}|(S_{1,4,2} \times S_{3,3,3} \times S_{4,3,4})^{1/3} \\
&\quad + \ln|X_{10}^{r_1}|(S_{1,4,3} \times S_{3,3,3} \times S_{4,3,4})^{1/3} + \ln|X_{11}^{r_1}|(S_{1,4,3} \times S_{3,3,4} \times S_{4,3,3})^{1/3} \\
&\quad + \ln|X_{12}^{r_1}|(S_{1,4,3} \times S_{3,3,4} \times S_{4,3,4})^{1/3} + \ln|X_{14}^{r_1}|(S_{1,4,4} \times S_{3,3,4} \times S_{4,3,3})^{1/3} \\
&\quad + \ln|X_{15}^{r_1}|(S_{1,4,4} \times S_{3,3,4} \times S_{4,3,4})^{1/3} \\
&= \ln(1)(0.0000 \times 0.9111 \times 0.8555)^{1/3} \\
&\quad + \ln(1)(0.0000 \times 0.9111 \times 0.0221)^{1/3} \\
&\quad + \ln(5)(0.1282 \times 0.9111 \times 0.221)^{1/3} \\
&\quad + \ln(5)(0.1282 \times 0.0195 \times 0.8555)^{1/3}
\end{aligned}
$$

$$
\begin{aligned}
&+ \ln(11)(0.1282 \times 0.0195 \times 0.221)^{1/3} \\
&+ \ln(6)(0.8718 \times 0.0195 \times 0.8555)^{1/3} \\
&+ \ln(6)(0.8718 \times 0.0195 \times 0.0221)^{1/3} \\
= {} & 0.0000 \times 0.0000 + 0.0000 \times 0.0039 \\
&+ 1.6094 \times 0.1371 + 1.6094 \times 0.1289 \\
&+ 2.3979 \times 0.0381 + 1.7918 \times 0.2442 + 1.7918 \times 0.0722 \\
= {} & 0.0000 + 0.0000 + 0.2207 + 0.2075 + 0.0913 + 0.4376 + 0.1293 \\
= {} & 1.0864
\end{aligned}
$$

Normalizing these values 0.0000, 0.4953, and 1.0864 gives $[w\mathrm{DT}_{13,1}^{r_1}, w\mathrm{DT}_{13,2}^{r_1}, w\mathrm{DT}_{13,3}^{r_1}] = [0.0000, 0.3131, 0.6869]$, as reported in Table IV. As with the whole set of condition attributes, the $\overline{\mathrm{DT}_q^{r_1}}$ and $\overline{w\mathrm{DT}_q^{r_1}}$ vectors can be represented as simplex coordinates in a simplex plot (see Figure 4).

In Figure 4, the clustering of the $\overline{\mathrm{DT}_q^{r_1}}$ and $\overline{w\mathrm{DT}_q^{r_1}}$ vectors is similar to those presented in the respective simplex plots concerned with $\overline{\mathrm{DT}_q^{C}}$ and $\overline{w\mathrm{DT}_q^{C}}$. An interesting example of the relationship between the condition classes associated with C and r_1 is that of $X_5^{r_1}$ and its relatives X_6^C and X_9^C (see Tables III and IV). The term relative identifies X_6^C and X_9^C have the same descriptor values for the condition attributes in r_1 but are augmented with descriptor values from c_2. In the simplex plots in Figure 3, the simplex coordinates of the associated $\overline{\mathrm{DT}_6^{C}}$ and $\overline{\mathrm{DT}_9^{C}}$ ($\overline{w\mathrm{DT}_6^{C}}$ and $\overline{w\mathrm{DT}_9^{C}}$) vectors are near to each other, and, importantly, near the respective simplex coordinate of the $\overline{\mathrm{DT}_5^{r_1}}$ ($\overline{w\mathrm{DT}_5^{r_1}}$) vector in the simplex plots presented in Figure 4.

One further concern is about the ξ^{r_1} and $\xi_w^{r_1}$ measures of certainty in overall classification that can be found for the condition classes associated with r_1. These are found to be $\xi^{r_1} = 0.2248$ and $\xi_w^{r_1} = 0.1774$. For both C and r_1, the ξ_w^C and $\xi_w^{r_1}$

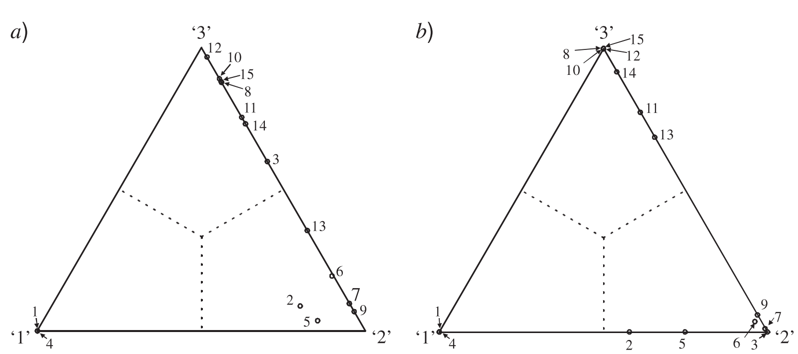

Figure 4. Simplex plot representation of the classification stability of each condition class.

values are less than their respective ξ^C and ξ^{r_1} values. This indicates that the use of relevance weighting has secured more certainty in classification of the condition classes. The exception is again with the $X_{13}^{r_1}$ condition class and its relative X_{26}^{C} (see Table IV), because the simplex coordinate of $X_{13}^{r_1}$ is nearer the incorrect "3" vertex than X_{26}^{C} previously was. The exposition here has been on the association of condition classes from subsets of condition attributes to the decision outcomes when a level of CVD is undertaken.

5. APPLICATION OF THE CONDITION CLASS STABILITY ANALYSIS WITH THE WINE DATA SET

To further illustrate the findings in this article on the effect of CVD, a slightly larger information system is briefly investigated. In this case the well-known wine data set is utilized.[18] Here, four condition attributes (of the 13) are utilized to characterize the 178 wines to one of three decision outcomes (wine cultivation approaches). The four condition attributes (and all 13) are continuous in nature; hence, without a level of CVD undertaken, a large level of granularity is inherent. Here, a naïve method of CVD is employed, which in this case is equal width CVD, on each of the four condition attributes. This is an unsupervised CVD technique that takes no account of the decision outcome values of each object (wine). For brevity, the details of the CVD process are included in the respective sets of estimated distributions describing the spread of the actual values in the intervals for each condition attribute, reported in Figure 5, found using the method of Parzen windows (as in Section 3).

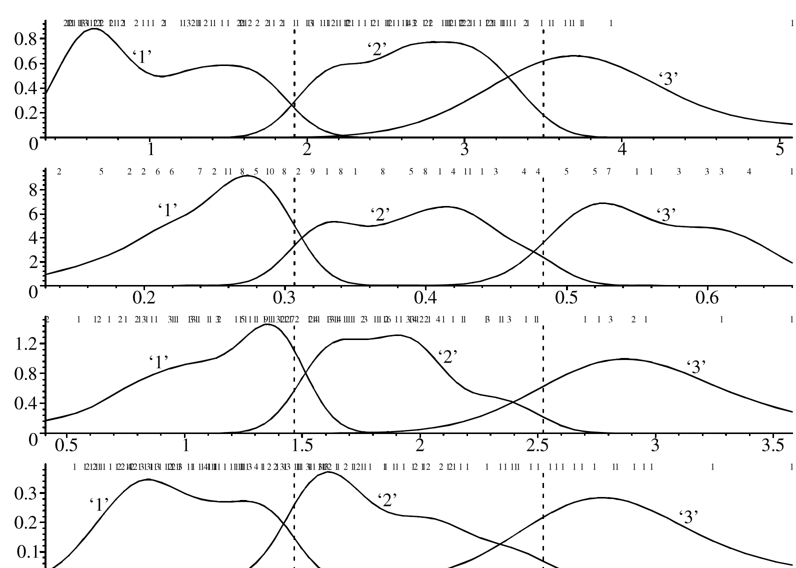

Figure 5. Sets of *pdf*s describing the four condition attributes in the wine data set.

Table V. $S_{n,j,r}$ values for the different intervals describing condition attributes.

$S_{n,j,r}$ value		Interval "1"	Interval "2"	Interval "3"	$S_{n,j,r}$ value		Interval "1"	Interval "2"	Interval "3"
c_1	"1"	**0.9702**	0.0298	0.0000	c_3	"1"	**0.8861**	0.1139	0.0000
	"2"	0.0374	**0.9414**	0.0213		"2"	0.0472	**0.9341**	0.0187
	"3"	0.0004	0.3221	**0.6775**		"3"	0.0000	0.1479	**0.8520**
c_2	"1"	**0.9206**	0.0792	0.0000	c_4	"1"	**0.9553**	0.0447	0.0000
	"2"	0.0553	**0.9009**	0.0438		"2"	0.1059	**0.8681**	0.0260
	"3"	0.0000	0.0766	**0.9234**		"3"	0.0000	0.2031	**0.7968**

In Figure 5, the four sets of estimated distributions again show (as in Section 4) the level of overlap between the possible intervalizations undertaken. The evidence for this overlap is found in the positions and frequencies of the individual attribute values, shown across the top of each set of estimated distributions. The subsequent series of $S_{n,j,r}$ likelihood values can be constructed (see Table V).

Based on the CVD of the four condition attributes, with RST a total of 36 condition classes were identified, of which 26 individually included objects to the same decision outcome. The $\overline{\mathrm{DT}_q^C}$ and $\overline{w\mathrm{DT}_q^C}$ vectors for these 26 condition classes are then constructed (see Table VI).

Table VI. Stability details on condition classes X_q^C, $q = 1, \ldots, 26$.

q	$E(D)$	$X_q^C : \lvert X_q^C \rvert$	$[\mathrm{DT}_{q,1}^C, \mathrm{DT}_{q,2}^C, \mathrm{DT}_{q,3}^C]$	$[w\mathrm{DT}_{q,1}^C, w\mathrm{DT}_{q,2}^C, w\mathrm{DT}_{q,3}^C]$
1	3	[1, 1, 1, 2] : 5	0.1318, 0.1297, **0.7384**	0.1261, 0.1363, **0.7376**
2	3	[1, 1, 1, 3] : 1	0.1050, 0.0178, **0.8773**	0.1121, 0.0208, **0.8670**
3	2	[1, 1, 2, 1] : 3	0.0817, **0.5995**, 0.3188	0.1094, **0.6364**, 0.2542
4	3	[1, 1, 2, 2] : 1	0.1345, 0.2711, **0.5945**	0.1914, 0.3061, **0.5025**
5	3	[1, 2, 1, 2] : 6	0.1274, 0.1713, **0.7013**	0.1144, 0.1558, **0.7298**
6	3	[1, 2, 1, 3] : 7	0.0986, 0.0228, **0.8785**	0.0871, 0.0204, **0.8925**
7	2	[1, 2, 2, 1] : 3	0.0723, **0.6469**, 0.2808	0.1111, **0.6612**, 0.2278
8	3	[1, 2, 2, 2] : 2	0.1205, 0.2965, **0.5830**	0.1883, 0.3082, **0.5034**
9	2	[1, 2, 3, 1] : 1	0.0604, **0.7755**, 0.1641	0.1209, **0.7651**, 0.1140
10	3	[1, 3, 1, 3] : 2	0.1069, 0.0312, **0.8619**	0.0568, 0.0299, **0.9133**
11	2	[1, 3, 2, 1] : 4	0.0568, **0.7030**, 0.2402	0.0698, **0.7979**, 0.1323
12	3	[1, 3, 2, 2] : 1	0.0935, 0.3181, **0.5884**	0.1337, 0.4201, **0.4461**
13	3	[1, 3, 2, 3] : 2	0.0781, 0.0457, **0.8762**	0.1203, 0.0650, **0.8147**
14	3	[1, 3, 3, 3] : 1	0.0455, 0.0536, **0.9008**	0.1798, 0.1046, **0.7156**
15	1	[2, 1, 1, 2] : 2	**0.4683**, 0.1146, 0.4171	**0.4640**, 0.0836, 0.4523
16	2	[2, 1, 3, 1] : 2	0.2807, **0.6580**, 0.0613	0.3111, **0.6510**, 0.0379
17	1	[2, 1, 3, 2] : 2	**0.5342**, 0.3442, 0.1216	**0.5860**, 0.3372, 0.0768
18	1	[2, 2, 2, 2] : 7	**0.4098**, 0.2937, 0.2965	**0.5249**, 0.2276, 0.2475
19	2	[2, 2, 3, 1] : 2	0.2151, **0.7158**, 0.0691	0.2806, **0.6725**, 0.0470
20	2	[2, 3, 1, 1] : 1	0.2102, **0.6863**, 0.1035	0.1968, **0.6367**, 0.1664
21	1	[2, 3, 1, 2] : 1	**0.3718**, 0.3337, 0.2945	**0.2966**, 0.2637, 0.4397
22	2	[2, 3, 2, 1] : 2	0.1523, **0.7360**, 0.1117	0.2049, **0.7226**, 0.0725
23	2	[3, 1, 2, 1] : 1	0.3975, **0.5498**, 0.0527	0.5698, **0.3595**, 0.0706
24	1	[3, 1, 2, 2] : 2	**0.6534**, 0.2484, 0.0982	**0.7613**, 0.1320, 0.1067
25	1	[3, 1, 3, 2] : 1	**0.6966**, 0.2652, 0.0382	**0.6990**, 0.2689, 0.0321
26	1	[3, 2, 3, 2] : 1	**0.6407**, 0.2956, 0.0638	**0.6413**, 0.3066, 0.0521

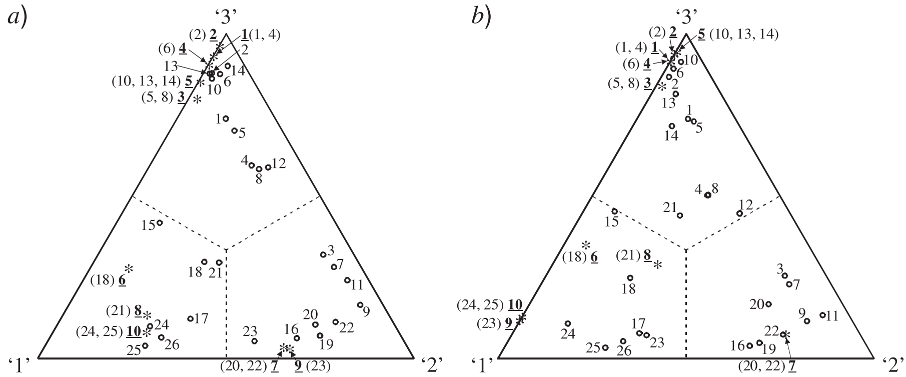

Figure 6. Simplex plot representation of the classification stability of each condition class.

An inspection of the vectors constructed shows that only two $\overline{w\mathrm{DT}_q^C}$ vectors ($\overline{w\mathrm{DT}_{21}^C}$ and $\overline{w\mathrm{DT}_{23}^C}$) include different values in bold and underlined. Because the wine data set is a three decision outcome problem, these vectors can also be represented in a series of simplex plots (see Figure 6).

In Figure 6 the individual $\overline{\mathrm{DT}_q^C}$ and $\overline{w\mathrm{DT}_q^C}$ vectors (shown as circles, with condition class index) are considerably more spread out across the whole of the respective simplex plot than in Figures 3 and 4, associated with the Iris data set. This shows that their simplex coordinates are farther away from the edges of the simplex plot, indicating that the uncertainty in their classification is to more than just two decision outcomes. Also shown in Figure 6 are the simplex coordinates of the $\overline{\mathrm{DT}_q^{r_1}}$ and $\overline{w\mathrm{DT}_q^{r_1}}$ vectors of a "near" reduct r_1 associated with C, namely $r_1 = \{c_1, c_2, c_4\}$, for which the specific details of the vectors (and condition classes) are reported in Table VII. In this case, these vectors are denoted by asterisks, with the condition class index shown in bold and underlined. Also included in brackets are the X_q^C relatives of each $X_q^{r_1}$ condition class.

Inspection of Table VII allows the further exposition of the simplex plots and the relationship between condition classes associated with different subsets of

Table VII. Stability details on condition classes $X_q^{r_1}$, $q = 1, \ldots, 10$.

q	$E(D)$	$X_q^{r_1} : \lvert X_q^{r_1} \rvert$	$[\mathrm{DT}_{q,1}^{r_1}, \mathrm{DT}_{q,2}^{r_1}, \mathrm{DT}_{q,3}^{r_1}]$	$[w\mathrm{DT}_{q,1}^{r_1}, w\mathrm{DT}_{q,2}^{r_1}, w\mathrm{DT}_{q,3}^{r_1}]$	X_q^C
1	3	[1, 1, 2] : 6	0.0682, 0.0000, **0.9318**	0.0829, 0.0000, **0.9171**	1, 4
2	3	[1, 1, 3] : 1	0.0359, 0.0000, **0.9641**	0.0621, 0.0000, **0.9379**	2
3	3	[1, 2, 2] : 8	0.1763, 0.0234, **0.8003**	0.1436, 0.0147, **0.8417**	5, 8
4	3	[1, 2, 3] : 7	0.0929, 0.0012, **0.9059**	0.0824, 0.0008, **0.9168**	6
5	3	[1, 3, 3] : 5	0.1421, 0.0047, **0.8532**	0.0525, 0.0032, **0.9443**	10, 13, 14
6	1	[2, 2, 2] : 7	**0.6217**, 0.0995, 0.2788	**0.5933**, 0.0570, 0.3496	18
7	2	[2, 3, 1] : 3	0.3294, **0.6365**, 0.0342	0.2018, **0.7246**, 0.0736	20, 22
8	1	[2, 3, 2] : 1	**0.6431**, 0.2221, 0.1348	**0.4323**, 0.2775, 0.2902	21
9	2	[3, 1, 1] : 1	0.3168, **0.6538**, 0.0294	0.8830, **0.0000**, 0.1170	23
10	1	[3, 1, 2] : 3	**0.6709**, 0.2475, 0.0816	**0.8729**, 0.0000, 0.1271	24, 25

condition attributes. For example, the cluster of $X_{10}^{r_1}$ and its relatives X_{24}^{C} and X_{25}^{C} in Figure 6a (near "1" vertex). With respect to the overall levels of certainty in the classification of the condition classes, the respective distance values are considered. For the four condition attributes in C, $\xi^{C} = 0.4184$ and $\xi_{w}^{C} = 0.4403$; also for the $r_1 = \{c_1, c_2, c_4\}$, $\xi^{r_1} = 0.3055$ and $\xi_{w}^{r_1} = 0.3716$. These sets of values show the $\xi_{w}^{\bullet}$ values to be consistently more than the respective $\xi^{\bullet}$ values. This is in contrast to the values in the Iris data set (see Section 4). One reason for these differences may be in the actual condition classes associated with C, which are not present in the condition classes associated with r_1, a possibility left for future research.

6. CONCLUSIONS

This article has investigated the effect of continuous value discretization (CVD) on the strength of the known classification of a condition class to a single decision outcome. Central to this notion is the modeling of the distribution of the values in each of the intervals constructed from the CVD undertaken. This distribution allows levels of likelihood on the association of the objects inside an interval to that interval and the other possible intervals constructed. The notion of likelihood of association is further considered at the more general condition class level with the possibility of the associated classification of condition classes to their known decision outcomes being uncertain. These findings could in future studies be used to allow further measurement of the effectiveness of the CVD undertaken.

The vectorization of the association of condition classes allows the introduction of the notion of the condition class space. In the case of three decision outcomes, the simplex plot representation allows a visual understanding of the classification stability. This is taken further with the relationship between condition classes from different subsets of condition attributes. The ability to quantify the level of overall certainty in the classification of objects is shown. Although a number of "near" reducts are considered (from RST), the role of the findings in this study in reduct selection still needs to be formally exposited. Whether these measures aid in the selection of reducts in RST will depend on the acknowledgement of "inexactness" inherent with CVD.

References

1. Pawlak Z. Rough sets. Int J Inform Comput Sci 1982;11:341–356.
2. Pawlak Z. Rough sets—Theoretical aspects of reasoning about data. London: Kluwer Academic Publishers; 1991.
3. Teghem J, Benjelloun M. Some experiments to compare rough sets and ordinal statistical methods. In: Słowiński R, editor. Intelligent decision support: Handbook of applications and advances of rough set theory, vol 11. Dordrecht: Kluwer; 1992. pp 267–284.
4. Dünstch I, Gediga G. Statistical evaluation of rough set dependency analysis. Int J Hum Comput Stud 1997;46:589–604.
5. Domingos P. The role of Occam's razor in knowledge discovery. Data Mining Knowl Discov 1999;3:409–425.

6. Bazan TG. A comparison of dynamic and non-dynamic rough sets methods for extracting laws from decision tables. In: Polkowski L, Skowron A, editors. Rough sets in knowledge discovery 1: Methodology and applications. New York: Physica-Verlag; 1998. pp 451–482.
7. Nguyen HS. Discretization problem for rough sets methods. In: Polkowski L, Skowron A, editors. Rough sets and current trends in computing. Proc First Int Conf RSCTC'98, Warsaw, Poland; 1998. pp 545–552.
8. Wright GN, Bolger FMI. Expertise and decision support. New York: Plenum; 1992.
9. Dougherty J, Kohavi R, Sahami M. Supervised and unsupervised discretization of continuous features. In: Proc 12th Int Conf on Machine Learning, Tahoe City. San Francisco: Kaufmann; 1995. pp 194–202.
10. Liu H, Setiono R. Dimensionality reduction via discretization. Knowl Base Syst 1996;9:67–72.
11. Beynon MJ, Peel MJ. Variable precision rough set theory and data discretisation: An application to corporate failure prediction. Omega 2001;29:561–576.
12. Chmielewski M, Grzymala-Busse L. Global discretization of continuous attributes as preprocessing for machine learning. Int J Approx Reas 1996;15:319–331.
13. Nguyen HS, Nguyen SH. Discretisation methods in data mining. In: Polkowski L, Skowron A, editors. Rough sets in knowledge discovery 1: Methodology and applications. New York: Physica-Verlag; 1998. pp 451–482.
14. Beynon MJ. Stability of continuous value discretisation: An application within rough set theory. Int J Approx Reas 2004;35:29–53.
15. Słowiński K, Słowiński R. Sensitivity analysis of rough classification. Int J Man Mach Stud 1990;32:693–705.
16. Browne C, Dünstch L, Gediga G. IRIS revisited: A comparison of discriminant and enhanced rough set data analysis. In: Polkowski L, Skowron A, editors. Rough sets in knowledge discovery 2: Applications, case studies and software systems. New York: Physica-Verlag; 1998. pp 345–368.
17. Parzen E. On estimation of a probability density function mode. Ann Math Stat 1962;33:1065–1076.
18. Forina M, Leardi R, Armanino C, Lanteri S. PARVUS: An extendable package of programs for data exploration, classification and correlation. Amsterdam: Elsevier; 1988.