Rule Discovery Based on New Attributes Construction

Agnieszka Dardzińska-Głębocka [1,2]

Department of Mathematics, Białystok Technical University
ul. Wiejska 45A, 15-351 Białystok, Poland

Abstract

This paper presents a method of constructing new attributes as a linear combination of original ones. Decision table based on n classification attributes and containing k-objects is seen in this paper as a collection of k points in n-dimensional space. For simplicity reason, it is assumed that the decision attribute is a binary one and the objects are partitioned into positive and negative. The problem is to find an efficient procedure for constructing possibly the smallest number of hyperplanes so each area surrounded by them only contains either positive or negative points. What is new in this paper is a strategy used to construct such hyperplanes. The work suggests unified approach to determine such attributes and use them for discovering new, more effective rules in decision systems.

Key words: Decision systems, knowledge mining, artificial attributes.

1 Introduction

Good data representations are often crucial for solving problems in intelligent systems area. Finding optimal or semi-optimal data representations for solving a given problem can be difficult and time consuming task. This is also true in data mining unless positive and negative examples are presented in a reasonably good form for knowledge discovery. One possible solution, to overcome such problems, is to build decision systems with some adaptive features. For instance, new attributes which are more appropriate for knowledge discovery, for a given data, can be generated by the system. Presented paper shows how to construct new attributes from the attributes available in the decision system, and then how to extract rules based on such attributes. These

[1] The author is grateful to prof. Zbigniew Ras for his useful remarks and suggestions which helped to improve this paper.
[2] Email: adardzin@uncc.edu

new rules usually have higher confidence and support than the rules extracted from original attributes. Before going any further into discussion of new problems, construction of such attributes is described. When constructed, they are treated exactly the same way as the original ones while they are used to discover rules.

2 Decision Systems

This section starts with the definition of an information system and a decision system. Next, the notion of a rule, its support and confidence is recalled.

Definition 2.1 By an *information system* we mean a triple $S = (X, A, V)$, where X is a nonempty, finite set of objects, A is a nonempty finite set of attributes, and $V = \bigcup\{V_a : a \in A\}$ is a set of their values. We assume that:

- V_a, V_b are disjoint for any $a, b \in A$ such that $a \neq b$,
- $a : X \longrightarrow V_a$ is a function for every $a \in A$.

Information systems can be seen as generalizations of decision tables. In any decision table together with the set of attributes a partition of that set into conditions and decisions is given. For simplicity reason, we consider decision tables with only one decision. Therefore a definition of a decision table is formed as follows.

Definition 2.2 By a *decision system* we mean any information system of the form $S = (X, A \cup \{d\}, V)$, where $d \notin A$ is a distinguished attribute called the *decision*. Attributes in A are called *classification* attributes.

Definition 2.3 By a *set of terms* for S we mean a least set T such that:

- $\mathbf{0}, \mathbf{1} \in T$,
- $w \in T$, $\sim w \in T$ for any $w \in V$,
- if $t_1, t_2 \in T$, then $(t_1 + t_2), (t_1 * t_2) \in T$.

Definition 2.4 Term t is called *simple* if $t = t_1 * t_2 * ... * t_n$ and $(\forall j \in \{1, 2, ..., n\})[(t_j \in V) \vee (t_j = \sim w \wedge w \in V)]$.

Definition 2.5 *Semantics* M of terms in S is defined in a standard way as follows:

- $M(\mathbf{0}) = \emptyset$, $M(\mathbf{1}) = X$,
- $M(w) = \{x \in X : w = a(x)\}$ for any $w \in V_a$,
- $M(\sim w) = X - M(w)$ for any $w \in V_a$,
- if t_1, t_2 are terms, then
  $M(t_1 + t_2) = M(t_1) \cup M(t_2)$,
  $M(t_1 * t_2) = M(t_1) \cap M(t_2)$,

Table 1
Decision System

X	a	b	d
x_1	60	60	-
x_2	80	60	+
x_3	100	60	+
x_4	130	60	-
x_5	60	50	+
x_6	130	50	+
x_7	80	40	-
x_8	100	40	+
x_9	130	40	-
$x_1$0	100	30	-

Definition 2.6 By a *rule* in a decision system S we mean any structure of the form $t \longrightarrow d_1$, where t is a simple term for S and $d_1 \in Dom(d)$.

Definition 2.7 *Support* of a rule $t \longrightarrow d_1$ (denoted as $sup(t \longrightarrow d_1)$ is defined as $sup(t * d_1)$ which means the number of objects in S having property $t * d_1$.

Definition 2.8 *Confidence* of a rule $t \longrightarrow d_1$ (denoted as $conf(t \longrightarrow d_1)$ is defined as $sup(t \longrightarrow d_1)/sup(t)$.

In this paper we assume that all classification attributes in S are numerical. For simplicity reason, and clear explanation, only two attributes a and b are taken into consideration in the example below.

Example 2.9 Assume that $S = (X, A, V)$ is the information system (Tab.1.), where: $X = \{x_1, x_2, ..., x_{10}\}$, $A = \{a, b\} \cup \{d\}$, and domains of attributes are respectively as follows: $dom(a) = [0, 200]$, $dom(b) = [0, 100]$, $dom(d) = \{+, -\}$.

The abov e decision system can be represented as a collection of 10 points in 2-dimensional space. These points are partitioned into t wo classes representing 2 v alues of attribute d. Their graphical representation is given in Fig. 1

Now, applying system LERS to S (see [2]), such certain rules are obtained:

$(b, 50) \longrightarrow (d, +)$ $(b, 30) \longrightarrow (d, -)$
$(a, 100) * (b, 60) \longrightarrow (d, +)$ $(a, 60) * (b, 60) \longrightarrow (d, -)$
$(a, 100) * (b, 40) \longrightarrow (d, +)$ $(a, 80) * (b, 40) \longrightarrow (d, -)$
$(a, 80) * (b, 60) \longrightarrow (d, +)$ $(a, 130) * (b, 40) \longrightarrow (d, -)$
$(a, 130) * (b, 60) \longrightarrow (d, -)$

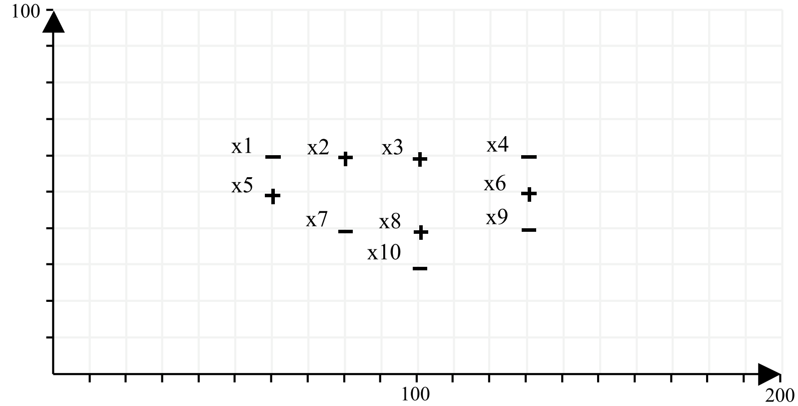

Fig. 1. Graphical interpretation of decision attributes

These rules have very small support which is due to the unpleasant distribution of 10 points which does not support construction of rectangles, parallel to axes a and b, with all points belonging to the same decision class. Slezak and Wroblewski proposed to handle this problem through introduction of new attributes being linear combinations of existing ones (see [5]). Their line of thought is continued in [1]. However, the authors did not give a method to achieve that goal. Saeed and Dardzinska in [4] proposed a strategy for classification and automatic identification of Arabic characters. Their strategy is based on construction of new attributes with values being angles between certain lines and distances between certain points. Similar idea is suggested in this paper.

3 Construction of new attributes

Before presenting a method for constructing new attributes and their domains, let us go back to the example given in the previous section. For simplicity reason needed in this section, this example only co v ersdecision tables with two classification attributes.

3.1 Analysis and interpretation

First of all, between two attributes a and b, the one with the biggest span is chosen. In our example, it is the attribute a illustrated on horizontal axis (Fig. 2). Its two boundary values are: $(0,0)$, $(200,0)$.

Then for these two boundary values we calculate the distance between them and a point representing one of the objects in S (values d_1, d_2 are calculated for object x_5). Also we calculate the angles between lines linking this object with these two boundary values and the axis representing attribute a. This step is repeated for the rest of the points representing objects in S. All these new quantities are used to form new decision table (Tab. 2) with new values describing objects from X.

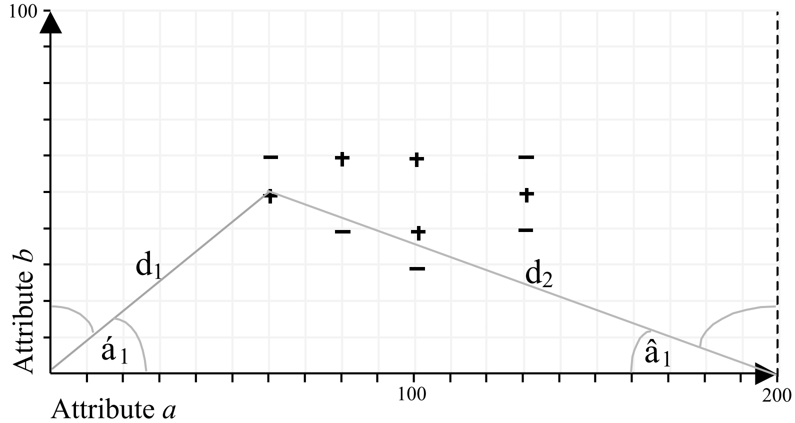

Fig. 2. Method of finding characteristic features

Table 2
Decision System

X	d_1	α_1	d_2	β_1	d
x_1	84	45	153	23	$-$
x_2	100	36	130	26	$+$
x_3	129	31	129	31	$+$
x_4	145	25	90	42	$-$
x_5	80	40	150	20	$+$
x_6	140	20	89	35	$+$
x_7	89	26	128	18	$-$
x_8	105	21	105	21	$+$
x_9	135	17	82	31	$-$
$x_1$0	100	16	100	16	$-$

These four new attributes and their values make possible to describe bounded areas representing quadrilaterals instead of rectangles (Fig. 3).

Because of this property, the following optimal rule with support 5 can be extracted:

$(\alpha_1, 20\ldots40) * (\beta_1, 20\ldots35) - \rightarrow (d,+)$.

Two discovered rules with support 2 are listed below:

$(\alpha_1, 16\ldots17) - \rightarrow (d, -)$.
$(\beta_1, 16\ldots18) - \rightarrow (d, -)$.

Notation $(\alpha, 20\ldots40)$ represents here the area covered by α and bounded by angles 20 and 40. To present the most general scenario, let us assume

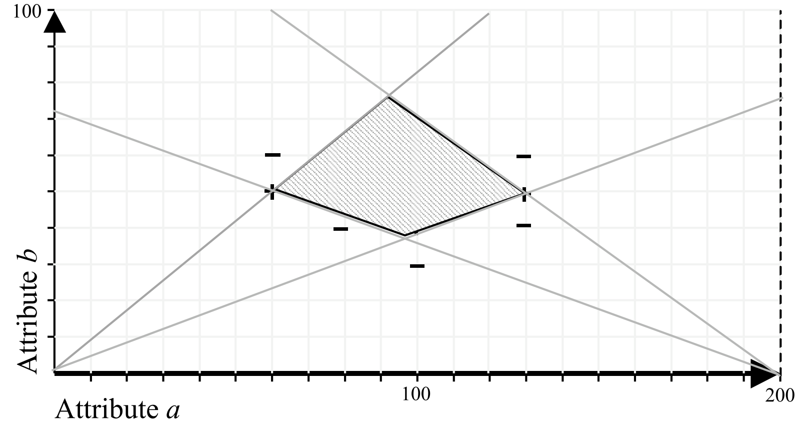

Fig. 3. Quadrilateral corresponding to the classification part of a rule describing concept "+".

that a decision table $S = (X, A \cup \{d\}, V)$ with $V = \bigcup\{V_{aj} : aj \in A\}$ has k classification attributes ($A = \{a_1, a_2, \ldots, a_k\}$) and that $max(V_{aj})$ denotes the maximal element and $min(V_{aj})$ the minimal element in V_{aj}. The span of an attribute a_j is defined as the value $max(V_{aj}) - min(V_{aj})$. Attribute with a minimal span is identified and used as a leading one in the process of new attributes construction. Namely, if a_m is an attribute with a minimal span then for any object

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{i(m-1)}, x_{i(m+k)}, \ldots, x_k) \in X,$$

$(k + 1)$ new objects are constructed where:

$$y_i = (y_{i1}, y_{i2}, \ldots, y_{i(m-1)}, y_{i(m+k)}, \ldots, y_k),$$
$$z_i = (z_{i1}, z_{i2}, \ldots, z_{i(m-1)}, z_{i(m+k)}, \ldots, z_k), \text{ and}$$
$$z_{ij} = x_{ij} - min(V_{aj}) \text{ for any } j \in \{1, 2, \ldots, m-1, m, m+1, \ldots, k\}.$$

T osimplify the notation, assume that $w_{ij} = (0, \ldots, 0, z_{ij}, 0, \ldots, 0)$ is a point on the axis representing attribute a_j and $[w_{ij}, z_i]$ denotes the vector from the point w_{ij} to z_i. Similarly, $[w_{ij}, y_i]$ denotes the v ector from the point w_{ij} to y_j and $[0, z_i]$, $[0, y_i]$ denote v ectors from the origin to z_i, y_i, respectively .
New attributes α_i, β_{ij}, d_i, d_{ij} for $j \in \{1, 2, \ldots, k\} - \{m\}$ are defined as:

$$\alpha_i = \angle([0, z_i], [0, y_i])$$
$$\beta_{ij} = \angle([w_{ij}, z_i], [w_{ij}, y_i]),$$
$$d_i = |[0, z_i]|, \; d_{ij} = |[w_{ij}, z_i]|$$

where $|\alpha|$ denotes the length of vector α and $\angle(\alpha, \beta)$ denotes the angle between v ectors α, β.
This wa y any object x_i of $k-$coordinates is replaced b y a new object of $2k-$coordinates, so the knowledge about attribute d can now be determined in terms of these new attributes.

3.2 Rule optimization method.

In this section, rule optimization method based on changing the boundary area of the quadrilateral describing the classification part of a rule is proposed. For simplicity reason, it will be described only for 2−dimension scenario but its general case is quite similar.
Assume that

$$r = [\alpha, \alpha_1 \ldots \alpha_2] * [\beta, \beta_1 \ldots \beta_2] - \rightarrow +$$

is a rule and $\{x_i\}_{i \in I}$ are objects supporting r. Also, assume that:

$$\alpha(x_{i1}) = \alpha_1,\ \alpha(x_{i2}) = \alpha_2,\ \alpha(x_{i3}) = \beta_1,\ \alpha(x_{i4}) = \beta_2$$

which means that $x_i 1, \ldots x_i 4$ are boundary objects for the area covered by the term

$$[\alpha, \alpha_1 \ldots \alpha_2] * [\beta, \beta_1 \ldots \beta_2].$$

Let's take first the object $x_{i1} = (a_{i1}, b_{i1})$ which lies on a boundary line of the area described by the condition part of the rule r. It is a line connecting two characteristic points: one is the origin $(0, 0)$, the second one is the object x_{i1} (see Fig. 4a). The point x_{i1} establishes contact, and the other characteristic point, which is $(0, 0)$, can change one of its coordinates taking a new value either within $[(0, 0), (a_{i1}, 0)]$ interval or $[(0, 0), (0, b_{i1})]$ interval (see Fig. 4b). The goal here is to replace the characteristic point $(0, 0)$ by a new one, so the new line crossing this new point and the point x_{i1} is also crossing one of the positive objects in the area covered by the term $[\alpha, \alpha_1 \ldots \alpha_2] * [\beta, \beta_1 \ldots \beta_2]$. Also, the area between the new line replacing the line corresponding to angle α_1 and between the other three lines (corresponding to angles α_2, β_1, β_2) have to cover the same positive objects as the objects covered by $[\alpha, \alpha_1 \ldots \alpha_2] * [\beta, \beta_1 \ldots \beta_2]$ and can not cover any negative objects. The algorithm implementing the construction of such a new line is quite simple and its complexity is linear from the point of view of the number of positive objects covered by the term $[\alpha, \alpha_1 \ldots \alpha_2] * [\beta, \beta_1 \ldots \beta_2]$. The new line replacing the line corresponding to angle α is shown in Fig. 4b.

Let's now take the object $x_{i3} = (a_{i3}, b_{i3})$. which lies on one of the boundary lines of the area described by the condition part of the rule r. It is a line connecting two characteristic points: one is the point $(0, w_a)$, the second one is the object x_{i3}. The point x_{i3} establishes contact and the other characteristic point, which is $(0, w_a)$, can change one of its coordinates taking a new value either within $[(a_{i3}, 0), (w_a, 0)]$ interval or $[(w_a, 0), (w_a, b_{i3})]$ interval (see Fig. 4c). The goal here is to replace the characteristic point $(w_a, 0)$) by a new one, so the new line crossing this new point and the point x_{i3} is also crossing one of the positive objects in the area covered by the term $[\alpha, \alpha_1 \ldots \alpha_2] * [\beta, \beta_1 \ldots \beta_2]$. The next step of this process for object x_{i3} is the same one we followed for the x_{i1} object. Finally, we follow the same steps for objects x_{i2} and x_{i4}. This way a new boundary area is created which is strictly determined by objects

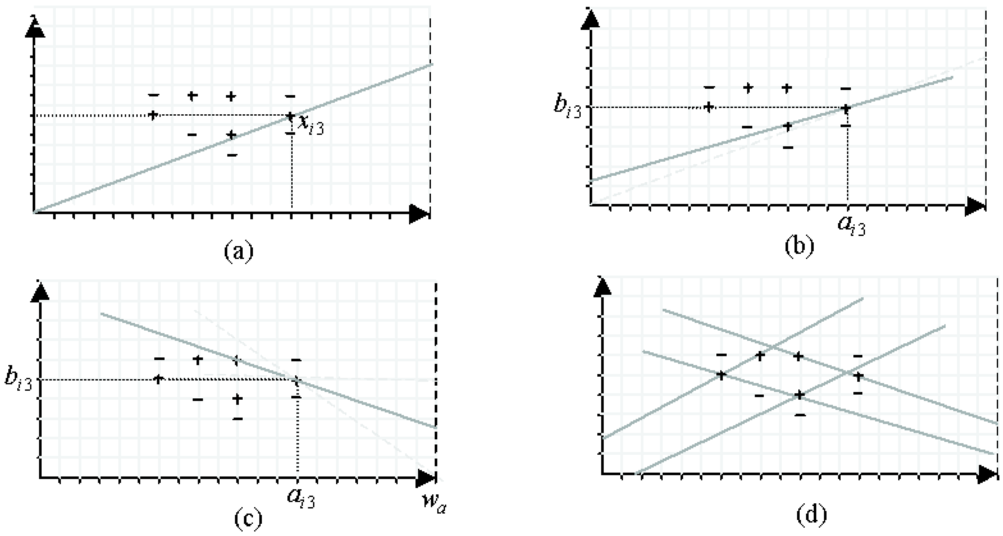

Fig. 4. Steps in finding new boundary area

classified in this example as positive (Fig. 4d).

4 Conclusion

This algorithm was initially tested on a database where either the confidence or support of discovered certain rules was rather low. Obtained results are quite promising for future work.

References

[1] Bazan, J., Szczuka, M., Wroblewski, J., "A new version of rough set exploration system", in *Proceedings of RSCTC'02 in Malvern, PA*, LNCS/LNAI, Springer-Verlag, No. 2475,2002, 397-404

[2] Grzymala-Busse, J. "A new version of the rule induction system LERS", in *Fundamenta Informaticae*, Vol. 31, No. 1, 1997,27-39

[3] Pawlak, Z., "Rough classification", in *International Journal of Man-Machine Studies*, Vol. 20, 1984, 469-483

[4] Saeed, K., Dardzinska, A., "Cursive letters language processing : Muqla model and Toeplitz matrices approach", in *Proceedings of the 4-th International Conference FQAS'2000*, Warsaw, Poland, October 25-28, 2000, Physica-Verlag, 326-333

[5] Slezak, D., Wroblewski, J., "Classification algorithms based on linear combinations of features", in *Proceedings of PKDD'99*, LNAI 1704, Springer-Verlag, 1999, 548-553