

Research Article

An Optimal Classification Method for Biological and Medical Data

Yao-Huei Huang,¹ Yu-Chien Ko,² and Hao-Chun Lu³

¹ *Institute of Information Management, National Chiao Tung University, Management Building 2, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan*

² *Department of Information Management, Chu-Hua University, No. 707, Section 2, WuFu Road, Hsinchu 300, Taiwan*

³ *Department of Information Management, College of Management, Fu Jen Catholic University, No. 510, Jhongjheng Road, Sinjhuang, Taipei 242, Taiwan*

Correspondence should be addressed to Yao-Huei Huang, yaohuei.huang@gmail.com

Received 25 October 2011; Revised 25 January 2012; Accepted 28 January 2012

Academic Editor: Jung-Fa Tsai

Copyright © 2012 Yao-Huei Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a union of hyperspheres by the mixed-integer nonlinear program to classify biological and medical datasets. A classifying program with nonlinear terms uses piecewise linearization technique to obtain a global optimum. The numerical examples illustrate that the proposed method can obtain the global optimum more effectively than current methods.

1. Introduction

Classification techniques have been widely applied in the biological and medical research domains [1–5]. Either objects classification or patterns recognition for biological and medical datasets necessarily demands an optimum accuracy for saving patients' lives. However, cancer identification with the supervised learning technique does not take a global view in identifying species or predicting survivals. The improvement should cover the whole scope to give implications instead of only considering the efficiency for diagnosis. This research aims to extract features from whole datasets in terms of induction rules.

In the given dataset with several objects, in which each object has some attributes and belongs to a specific class, classification techniques are used to find a rule of attributes that appropriately describes the features of a specified class. The techniques have been studied over the last four decades, including decision tree-based methods [6–11], hyperplane-based methods [12–14], and machine learning-based methods [14–17].

To assess the effects of these classifying techniques, three criteria are used for evaluating the quality of inducing rules based on the study of Li and Chen [3].

- (i) *Accuracy*. The rule fitting a class should not cover the objects of other classes. The accuracy of a rule should be the higher the better.
- (ii) *Support*. A good rule of fitting a class should be supported by most of the objects of the same class.
- (iii) *Compact*. A good rule should be expressed in a compact way. That is, the fewer the number of rules, the better the rules are.

This study proposes a novel method to induce rules with high rates of accuracy, support, and compactness based on global optimization techniques, which have become more and more useful in biological and medical researches.

The rest of this paper is organized as follows. Section 2 gives an overview of the related literatures. Two types of mathematical models and a classification algorithm are proposed in Section 3. The numerical examples demonstrate the effectiveness of the proposed method in Section 4. Finally, the main conclusions of this study and future work are drawn in Section 5.

2. Literature Review

Currently, two well-known methods are used to induce classification rules. The first method is the decision tree-based method, which has been developed in the last few decades [6–10]. It is widely applied to fault isolation of an induction motor [18] to classify normal or tumor tissues [19], skeletal maturity assessment [20], proteomic mass spectra classification [21], and other cases [22, 23]. Although the decision tree-based method assumes that all classes can be separated by linear operations, the inducing rules will suffer if the boundaries between the classes are nonlinear. In fact, the linearity assumption prohibits practical applications because many biological and medical datasets have complicated nonlinear interactions between attributes and predicted classes.

Consider the classification problem with two attributes as shown in Figure 1, where “○” represents a first-class object, and “●” represents a second-class object. Figure 1 depicts a situation in which a nonlinear relationship exists between the objects of two classes. Decision tree method focuses on inducing classification rules for the objects, as shown in Figure 1(b), in which the decision tree method requires four rectangular regions to classify the objects.

The second is the support vector hyperplane method, which conducts feature selection and rule extraction from the gene expression data of cancer tissue [24]; it is also applied in other applications [12–14, 25]. The technique separates observations of different classes by multiple hyperplanes. As the number of decision variables is required to express the relationship between each training datum and hyperplane, and the separating hyperplane is assumed a nonlinear programming problem, the training speed becomes slow for a large number of training data. Additionally, similar hypersphere support vector methods have been developed by Lin et al. [26], Wang et al. [27], Gu and Wu [28], and Hifi and M’Hallah [29] for classifying objects. In classification algorithms, they partition the sample space using the sphere-structured support vector machine [14, 30]. However, these methods need to form a classification problem as a nonlinear nonconvex program, which makes reaching an optimal solution difficult. Taking Figure 1 as an example, a hyperplane-based method requires four hyperplanes to discriminate the objects, as shown in Figure 2.

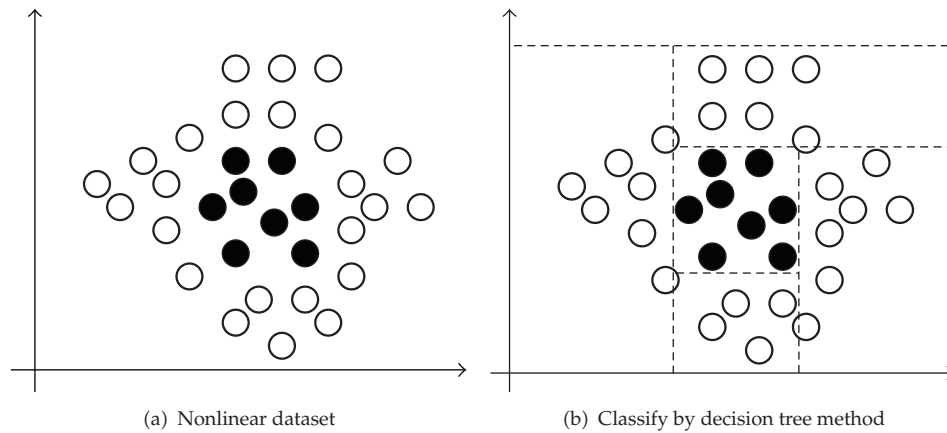


Figure 1: Classifying the objects of two classes.

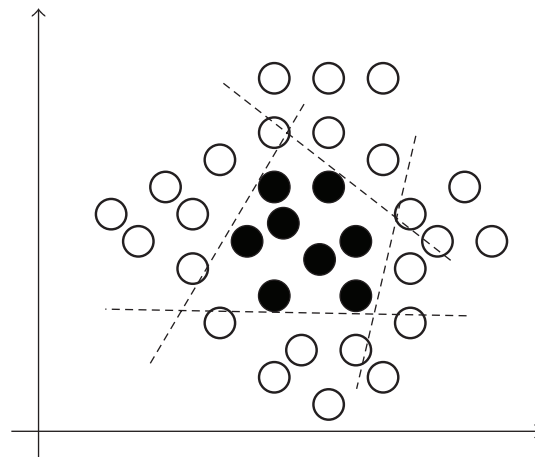


Figure 2: Classify by hyperplane method method.

As previously mentioned, many biological and medical datasets have complicated boundaries between attributes and classes. Both decision tree-based methods and hyperplane-based methods find only the rules with high accuracy, which either cover only a narrow part of the objects or require numerous attributes to explain a classification rule. Although these methods are computationally effective for deducing the classifications rules, they have two limitations as follows.

- (i) Decision tree-based methods are heuristic approaches that can only induce feasible rules. Moreover, decision tree-based methods split the data into hyperrectangular regions using a single variable, which may generate a large number of branches (i.e., low rates of compactness).
- (ii) Hyperplane-based methods use numerous hyperplanes to separate objects of different classes and divide the objects in a dataset into indistinct groups. The method may generate a large number of hyperplanes and associated rules with low rates of compactness.

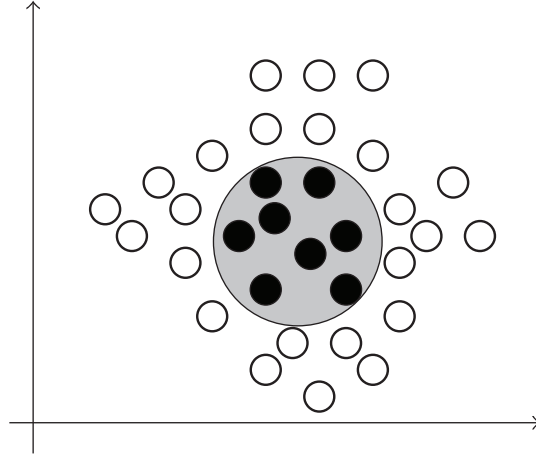


Figure 3: Classify by hypersphere.

Therefore, this study proposes a novel hypersphere method to induce classification rules based on a piecewise linearization technique. The technique reformulates the original hypersphere model by a piecewise linearization approach using a number of binary variables and constraints in the number of piecewise line segments. As the number of break points used in the linearization process increases, the error in linear approximation decreases, and an approximately global optimal solution of the hypersphere model can be obtained. That is, the proposed method is an optimization approach that can find the optimal rules with a high rate of accuracy, support, and compactness. The concept of the hypersphere method is depicted in Figure 3, in which only one circle is required to classify the objects. All objects of class “•” are covered by a circle, and those not covered by this circle belong to class “○.”

3. The Proposed Models and Algorithm

As the classification rules directly affect the rates of accuracy, support, and compactness, we formulate two models to determine the highest accuracy rate and support rate, respectively. To facilitate the discussion, the related notations are introduced first:

$a_{i,j}$: j 'th attribute value of the i 'th object,

$h_{t,k,j}$: j 'th center value of the k 'th hypersphere for class t ,

$r_{t,k}$: radius of the k 'th hypersphere for class t ,

$n(t)$: number of objects for class t ,

c_i : i 'th object belonging to class $c_i \in \{1, 2, \dots, g\}$,

m : number of attributes,

R_t : a rule describing class t .

Based on these notations, we propose two types of classification models as follow.

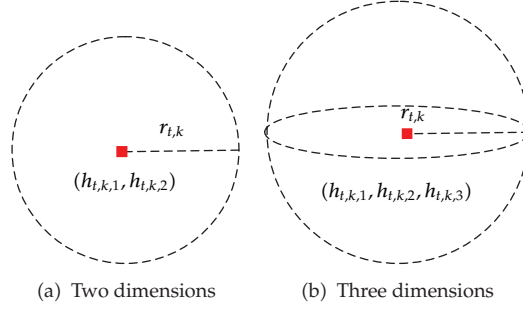


Figure 4: The concept of hypersphere method.

3.1. Two Types of Classification Models

Considering the object x_i and hypersphere $S_{t,k}$, and normalizing $a_{i,j}$ (i.e., to express its scale easily), we then have the following three notations.

Notation 1. Normalization rescales all $a_{i,j}$ as $a'_{i,j}$. The following is a normalizing formula:

$$a'_{i,j} = \frac{a_{i,j} - \underline{a}_j}{\bar{a}_j - \underline{a}_j}, \quad (3.1)$$

where $0 \leq a'_{i,j} \leq 1$, \bar{a}_j is the largest value of attribute j , and \underline{a}_j is the smallest value of attribute j .

Notation 2. A general form for expressing an object x_i is written as

$$x_i = (a'_{i,1}, a'_{i,2}, \dots, a'_{i,m}; c_i), \quad (3.2)$$

where c_i is the class index of object x_i .

Notation 3. A general form for expressing a hypersphere $S_{t,k}$ is written as

$$S_{t,k} = (h_{t,k,1}, h_{t,k,2}, \dots, h_{t,k,m}; r_{t,k}), \quad (3.3)$$

where $S_{t,k}$ is the k 'th hypersphere for class t .

We use two and three dimensions (i.e., two attributes and three attributes) as visualizations to depict clearly a circle and a sphere, respectively (Figure 4). Figure 4(a) denotes the centroid of the circle as $(h_{t,k,1}, h_{t,k,2})$ and the radius of the circle as $r_{t,k}$. They are extended to three dimensions called sphere (Figure 4(b)); in m dimensions (i.e., m attributes), $m > 3$, which are then called hyperspheres.

To find each center and the radius of the hypersphere, the following two nonlinear models are considered. The first model looks for a support rate as high as possible while the accuracy rate is fixed to 1, as shown in Model 1.

Model 1. One has the following:

$$\begin{aligned}
 & \text{Maximize } \sum_{i \in I^+} u_{t,i,k} \\
 & \text{subject to } \sum_{j=1}^m (a'_{i,j} - h_{t,k,j})^2 \leq (1 - u_{t,i,k})M + r_{t,k}^2, \quad \forall i \in I^+, \\
 & \sum_{j=1}^m (a'_{i',j} - h_{t,k,j})^2 > r_{t,k}^2, \quad \forall i' \in I^-, \\
 & u_{t,i,k} \in \{0, 1\} \quad \forall i \in I^+, r_{t,k} \geq 0, \text{ and } M \text{ is big enough constant,}
 \end{aligned} \tag{3.4}$$

where I^+ and I^- are the two sets for all objects expressed, respectively, by

$$I^+ = \{i \mid i = 1, 2, \dots, n, \text{ where object } i \in \text{class } t\}, \tag{3.5}$$

$$I^- = \{i' \mid i' = 1, 2, \dots, n, \text{ where object } i' \notin \text{class } t\}. \tag{3.6}$$

Referring to Li and Chen [3], the rates of accuracy and support of R_t in Model 1 can be specified by the following definitions.

Definition 3.1. The accuracy rate of a rule R_t for Model 1 is $AR(R_t) = 1$.

Definition 3.2. The support rate of a rule R_t for Model 1 is specified as follows.

- (i) If $\sum_{k \in K} u_{t,i,k} \geq 1$ for all i belonging to class t , then $U_{t,i} = 1$; otherwise $U_{t,i} = 0$, where K indicates the hypersphere set for class t .
- (ii)

$$SR(R_t) = \frac{\sum_{i \in \text{class } t} U_{t,i}}{n(t)}, \tag{3.7}$$

where $n(t)$ indicates the number of objects belonging to class t .

The second model looks for an accuracy rate as high as possible while the support rate is fixed to 1, as shown in Model 2.

Model 2. One has the following:

$$\begin{aligned}
 & \text{Maximize } \sum_{i' \in I^-} v_{t,i',k} \\
 & \text{subject to } \sum_{j=1}^m (a'_{i,j} - h_{t,k,j})^2 \leq r_{t,k}^2, \quad \forall i \in I^+,
 \end{aligned}$$

$$\sum_{j=1}^m (a'_{i',j} - h_{t,k,j})^2 > (v_{t,i',k} - 1)M + r_{t,k}^2, \quad \forall i' \in I^-,$$

$$v_{t,i',k} \in \{0, 1\}, \quad \forall i' \in I^-, \quad r_{t,k} \geq 0,$$
(3.8)

where I^+ and I^- are the two sets expressed by (3.5) and (3.6), respectively.

Similarly, the rates of accuracy and support of R_t in Model 2 can be considered as follows.

Definition 3.3. The accuracy rate of a rule R_t of Model 2 is denoted as $AR(R_t)$ and is specified as follows.

- (i) If $\sum_{k \in K} v_{t,i',k} = 0$ belongs to class t , then $V_{t,i'} = 1$ for all i' ; otherwise, $V_{t,i'} = 0$, where K represents the hypersphere set for class t .
- (ii)

$$AR(R_t) = \frac{\|R_t\| - \sum_{i' \in \text{class } t} V_{t,i'}}{\|R_t\|},$$
(3.9)

where $\|R_t\|$ represents the number of total objects covered by R_t .

Definition 3.4. The support rate of a rule R_t of Model 2 is denoted as $SR(R_t)$, and $SR(R_t) = 1$.

Definition 3.5. The compactness rate of a set of rules R_1, \dots, R_g , denoted as $CR(R_1, \dots, R_g)$, is expressed as follows:

$$CR(R_1, \dots, R_g) = \frac{g}{\sum_{t=1}^g US_t},$$
(3.10)

where US_t means the number of hyperspheres and unions of hyperspheres for class t . A union of hyperspheres indicates that the object is covered by different hyperspheres, as shown in Figure 5. Take Figure 5 for an example, in which there are two classes. The objects of class "○" are covered by two unions of the circles (i.e., $S_{1,1} \cup S_{1,2} \cup S_{1,3}$ and $S_{1,4} \cup S_{1,5}$), and the objects of class "●" are covered by one circle (i.e., $S_{2,1}$). Therefore, $US_1 = 2$, $US_2 = 1$, and $CR(R_1, R_2) = 2/3$.

Moreover, Models 1 and 2 are separable nonlinear programs solvable to find an optimal solution by linearizing the quadratic terms $h_{t,k,j}^2$. The piecewise linearization technique is discussed as follows.

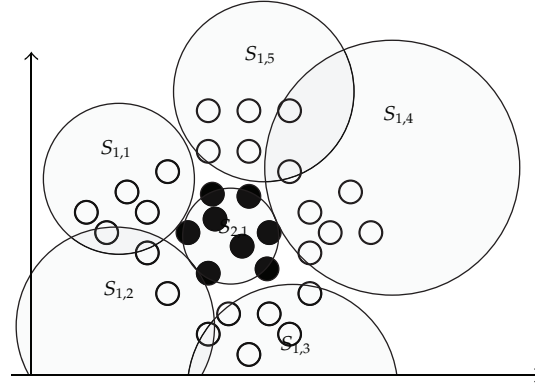


Figure 5: Classify by hypersphere method.

Proposition 3.6 (referring to Beale and Forrest [31]). Denote approximate function $L(f(x))$ as a piecewise linear function (i.e., linear convex combination) of $f(x)$, where $b_l, l = 1, 2, \dots, q$ represents the break points of $L(f(x))$. $L(f(x))$ is expressed as follows:

$$f(x) \cong L(f(x)) = \sum_{l=1}^q f(b_l) \omega_l, \quad (3.11)$$

$$x = \sum_{l=1}^q \omega_l b_l, \quad (3.12)$$

$$\sum_{l=1}^q \omega_l = 1, \quad (3.13)$$

where $\omega_l \geq 0$, and (3.13) is a special-ordered set of type 2 (SOS2) constraint (reference to Beale and Forrest [31]).

Note that the SOS2 constraint is a set of variables in which at most two variables may be nonzero. If two variables are nonzero, they must be adjacent in the set.

Notation 4. According to Proposition 3.6, let $f(x) = h_{t,k,j}^2$. $f(x)$ is linearized by the Proposition 3.6 and is expressed as $L(h_{t,k,j}^2)$.

3.2. Solution Algorithm

A proposed algorithm is also presented to seek the highest accuracy rate or the highest support rate, as described as follows.

Algorithm 3.7.

Step 1. Normalize all attributes (i.e., rescale $a'_{i,j} = (a_{i,j} - \underline{a}_j) / (\bar{a}_j - \underline{a}_j)$ to be $0 \leq a'_{i,j} \leq 1$).

Step 2. Initialization: $t = 1$ and $k = 1$.

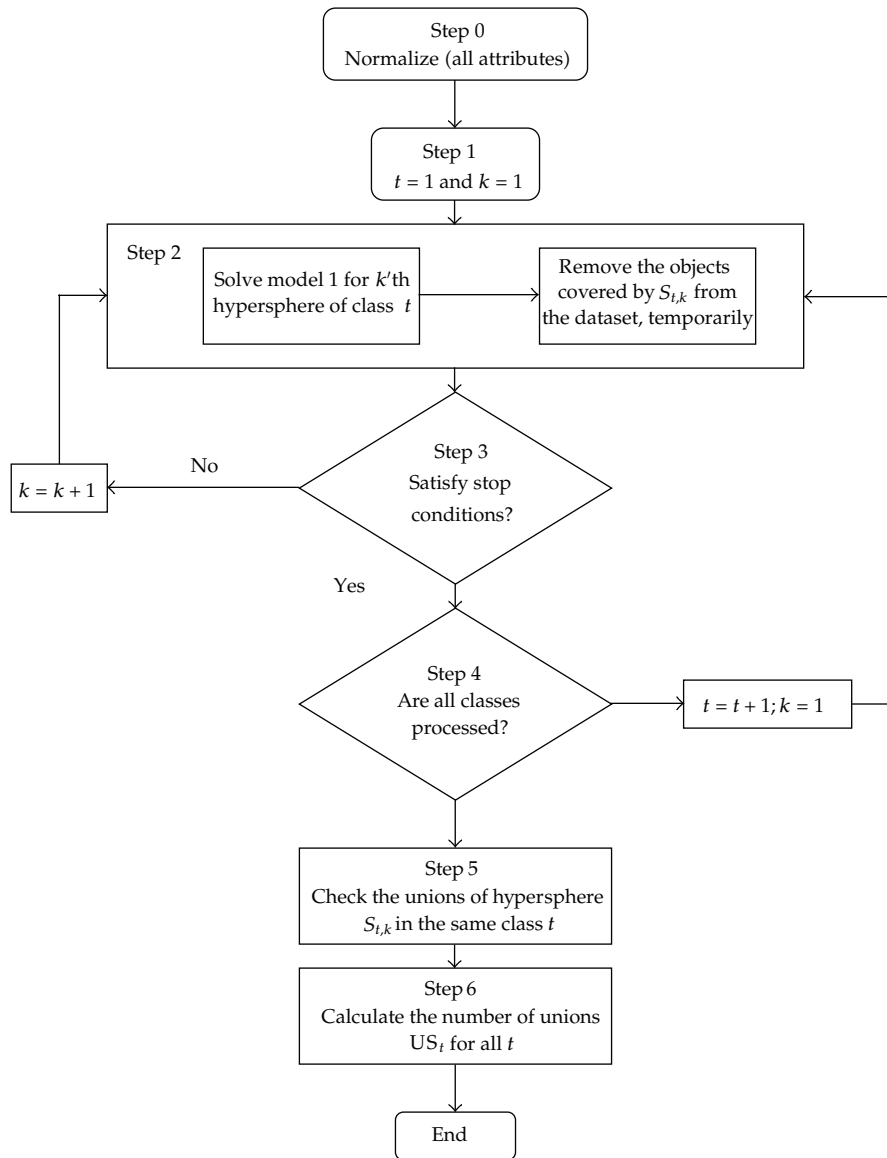


Figure 6: Flowchart of the proposed algorithm.

Step 3. Solve Model 1 (or Model 2) to obtain the k 'th hypersphere of class t . Remove the objects covered by $S_{t,k}$ from the dataset temporarily.

Step 4. Let $k = k + 1$, and resolve Model 1 (or Model 2) until all objects in class t are assigned to the hyperspheres of same class.

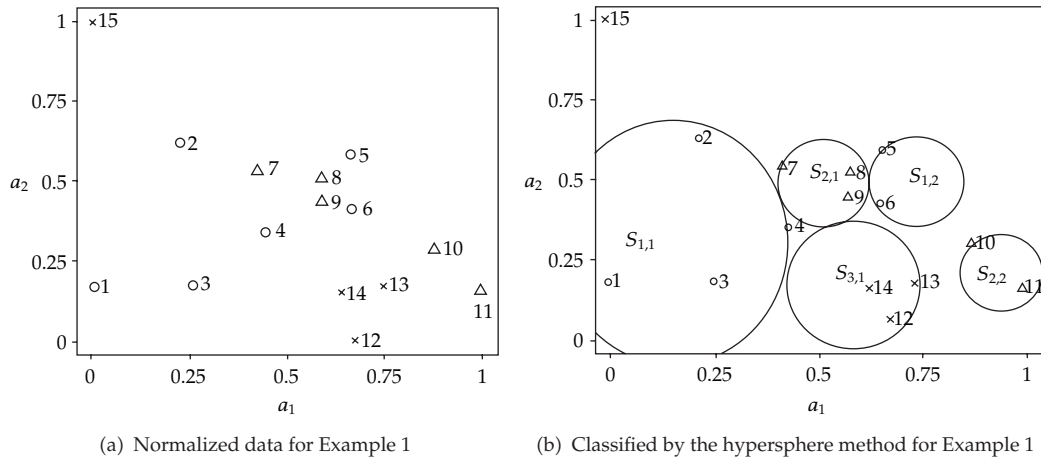
Step 5. Let $k = 1$ and $t = t + 1$, and reiterate Step 3 until all classes are processed.

Step 6. Check the independent hyperspheres and unions of hyperspheres $S_{t,k}$ in the same class t .

Table 1: Dataset of Example 1.

Object	$a_{i,1}$	$a_{i,2}$	c_i	Object	$a_{i,1}$	$a_{i,2}$	c_i	Object	$a_{i,1}$	$a_{i,2}$	c_i
x_1	6	8	1	x_2	12	20	1	x_3	13	8	1
x_4	18	12.5	1	x_5	24	19	1	x_6	24	14.5	1
x_7	17.5	17.5	2	x_8	22	17	2	x_9	22	15	2
x_{10}	30	11	2	x_{11}	33.5	7.5	2	x_{12}	24.5	3.5	3
x_{13}	26.5	8	3	x_{14}	23.5	7.5	3	x_{15}	6	30	3

(Symbols of c_i) 1: \circ 2: Δ 3: \times .

**Figure 7:** Visualization of Example 1.

Step 7. Calculate and record the number of independent hyperspheres and unions of hyperspheres in US_t , and iterate t until all classes are done.

According to this algorithm, we can obtain the optimal rules to classify objects most efficiently. The process of the algorithm is depicted in Figure 6.

3.3. Operation of a Simple Example

Consider a dataset T in Table 1 as an example, which has object i , two attributes $(a_{i,1}, a_{i,2})$, and an index of classes (c_i) for $i = 1, 2, \dots, 15$. The dataset T is expressed as $T = \{x_i \mid (a_{i,1}, a_{i,2}; c_i) \forall i = 1, 2, \dots, 15\}$. There are the domain values of $c_i \in \{1, 2, 3\}$. As there are only two attributes, these 15 objects can be plotted on a two-dimensional space after normalizing them, as shown in Figure 7(a).

This example can be solved by the proposed algorithm as follows.

Step 1. Normalize all attributes (i.e., $a'_{i,1} = (a_{i,1} - 6)/(33.5 - 6)$ and $a'_{i,2} = (a_{i,2} - 3.5)/(30 - 3.5)$).

Step 2. Initialization: $t = 1$ and $k = 1$.

Table 2: Centroid points for the Iris data set by the proposed method.

Rule number	Union of spheres	$S_{t,k}$	$h_{t,k,1}$	$h_{t,k,2}$	$h_{t,k,3}$	$h_{t,k,4}$	$r_{t,k}$
R_1	$S_{1,1}$	$S_{1,1}$	0.366	0.807	0.000	0.000	0.6205
R_2	$S_{2,1} \cup S_{2,2} \cup S_{2,3}$	$S_{2,1}$	0.557	0.320	0.205	0.460	0.2540
		$S_{2,2}$	0.575	0.581	0.626	0.515	0.0612
		$S_{2,3}$	0.423	0.261	0.352	0.490	0.1388
R_3	$S_{3,1} \cup S_{3,2}$	$S_{3,1}$	0.248	0.000	2.226	2.151	4.8087
		$S_{3,2}$	0.329	0.187	0.650	0.613	0.0330

Table 3: Comparing results for the Iris flower data set (R_1, R_2, R_3).

Items	Proposed method	Decision tree	Hyperplane support vector
$AR(R_1, R_2, R_3)$	(1,1,1)	(1,0.98,0.98)	(1,0.98,0.96)
$SR(R_1, R_2, R_3)$	(1,0.98,0.98)	(1,0.98,0.98)	(1,0.96,0.98)
CR	1	0.5	0.1875

Step 3. The classification model (i.e., Model 1) is linearly formulated as follows:

$$\begin{aligned}
 & \text{Maximize } \sum_{i \in I^+} u_{t,i,k} \\
 & \text{subject to } \sum_{j=1}^m \left(a'_{ij}{}^2 - 2a'_{ij} h_{t,k,j} + L(h_{t,k,j}^2) \right) \leq (1 - u_{t,i,k})M + r_{t,k}^2 \quad \forall i \in I^+, \\
 & \sum_{j=1}^m \left(a'_{i'j}{}^2 - 2a'_{i'j} h_{t,k,j} + L(h_{t,k,j}^2) \right) > r_{t,k}^2 \quad \forall i' \in I^-, \\
 & u_{t,i,k} \in \{0, 1\}, r_{t,k}^2 \geq 0,
 \end{aligned} \tag{3.14}$$

where $I^+ = \{x_1, x_2, \dots, x_6\}$ and $I^- = \{x_7, x_8, \dots, x_{15}\}$. The optimal solution of the $(h_{t,k,1}, h_{t,k,2}, r_{t,1}) = (0.047, 0.265, 0.15749)$ for $S_{1,1}$, where $S_{1,1}$ covers objects 1–4. We then temporarily remove these objects covered by $S_{1,1}$.

Step 4. $k = k + 1$: the optimal solution of the $(h_{t,k,1}, h_{t,k,2}, r_{t,k}) = (0.736, 0.5, 0.0138)$ for $S_{1,2}$, where $S_{1,2}$ covers objects 5–6. Class 1 is then done.

Step 5. As $t = t + 1$, $k = 1$, and Steps 3 and 4 are iterated, we then, respectively, have optimal solutions for $S_{t,k}$ as follows. The results are shown in Figure 7(b).

- (i) $(h_{2,1,1}, h_{2,1,2}, r_{2,1}) = (0.514, 0.469, 0.0127)$, where $S_{2,1}$ covers objects 7–9.
- (ii) $(h_{2,2,1}, h_{2,2,2}, r_{2,2}) = (0.929, 0.210, 0.0251)$, where $S_{2,2}$ covers objects 10–11.
- (iii) $(h_{3,1,1}, h_{3,1,2}, r_{3,1}) = (0.583, 0.188, 0.0436)$, where $S_{3,1}$ covers objects 12–14.

Step 6. Check and calculate the unions of hypersphere $S_{t,k}$ for all k in class t (i.e., Initial $t = 1$).

Step 7. As $t = t + 1$, mark the number of unions of class t into US_t and iterate Step 6 until $t = g$.

4. Numerical Examples

This study shows how the experimental results evaluate the performance, including accuracy, support, and compactness rates, and compares the proposed model with different methods using CPLEX [32]. All tests were run on a PC equipped with an Intel Pentium D 2.8 GHz CPU and 2 GMB RAM. Three datasets were tested in our experiments as follows:

- (i) Iris Flower dataset introduced by Sir Ronald Aylmer Fisher (1936),
- (ii) European barn swallow (*Hirundo rustica*) dataset obtained by trapping individual swallows in Stirlingshire, Scotland, between May and July 1997 [1, 3],
- (iii) the highly selective vagotomy (HSV) patient dataset of F. Raszeja Memorial Hospital in Poland [3, 33, 34].

4.1. Iris Flower Dataset

The Iris Flower dataset contains 150 objects. Each object is described by four attributes (i.e., sepal length, sepal width, petal length, and petal width) and is classified by one of three classes (i.e., setosa, versicolor, and virginica). By solving the proposed method, we induced six hyperspheres (i.e., $S_{1,1} \in \text{Class 1}$, $S_{2,1}, S_{2,2}, S_{2,3} \in \text{Class 2}$, and $S_{3,1}, S_{3,2} \in \text{Class 3}$). The induced classification rules are reported in Table 2. Table 2 also lists a hypersphere and two unions (i.e., $S_{1,1}$, $S_{2,1} \cup S_{2,2} \cup S_{2,3}$, and $S_{3,1} \cup S_{3,2}$) of hyperspheres with centroid points and radii.

Rule R_1 in Table 2 contains a hypersphere $S_{1,1}$, which implies that

- (i) "if $(a'_{i,1} - 0.366)^2 + (a'_{i,2} - 0.807)^2 + (a'_{i,3} - 0)^2 + (a'_{i,4} - 0)^2 \leq 0.6205$, then object x_i belongs to class 1."

Rule R_2 in Table 2 contains a union of three hyperspheres (i.e., $S_{2,1} \cup S_{2,2} \cup S_{2,3}$) which implies that

- (i) "if $(a'_{i,1} - 0.557)^2 + (a'_{i,2} - 0.32)^2 + (a'_{i,3} - 0.205)^2 + (a'_{i,4} - 0.46)^2 \leq 0.254$, then object x_i belongs to class 2," or
- (ii) "if $(a'_{i,1} - 0.575)^2 + (a'_{i,2} - 0.581)^2 + (a'_{i,3} - 0.626)^2 + (a'_{i,4} - 0.515)^2 \leq 0.0612$, then object x_i belongs to class 2," or
- (iii) "if $(a'_{i,1} - 0.423)^2 + (a'_{i,2} - 0.261)^2 + (a'_{i,3} - 0.352)^2 + (a'_{i,4} - 0.49)^2 \leq 0.1388$, then object x_i belongs to class 2."

Rule R_3 in Table 2 contains a union of two hyperspheres (i.e., $S_{3,1} \cup S_{3,2}$), which implies that

- (i) "if $(a'_{i,1} - 0.248)^2 + (a'_{i,2} - 0)^2 + (a'_{i,3} - 2.226)^2 + (a'_{i,4} - 2.151)^2 \leq 4.8087$, then object x_i belongs to class 3," or
- (ii) "if $(a'_{i,1} - 0.329)^2 + (a'_{i,2} - 0.187)^2 + (a'_{i,3} - 0.65)^2 + (a'_{i,4} - 0.613)^2 \leq 0.033$, then object x_i belongs to class 3."

Comparing the proposed method with both decision tree [3] and hyperplane methods [35] in deducing the classification rules for the Iris Flower dataset, Table 3 lists the experimental result.

Table 4: Centroid points for the Swallow data set by the proposed method.

Rule number	Union of spheres	$S_{t,k}$	$h_{t,k,1}$	$h_{t,k,2}$	$h_{t,k,3}$	$h_{t,k,4}$	$h_{t,k,5}$	$h_{t,k,6}$	$h_{t,k,7}$	$h_{t,k,8}$	$r_{t,k}$
R_1	$S_{1,1} \cup S_{1,2} \cup S_{1,3} \cup S_{1,4}$	$S_{1,1}$	0.607	0.110	0.806	0.000	0.000	0.406	1.077	1.163	1.780
		$S_{1,2}$	0.483	0.000	0.236	0.328	0.000	0.793	0.280	0.931	0.948
		$S_{1,3}$	0.588	0.000	1.179	0.000	0.000	0.653	1.037	0.982	1.879
		$S_{1,4}$	0.414	0.000	0.658	0.000	0.227	0.085	0.563	1.224	1.197
R_2	$S_{2,1} \cup S_{2,2} \cup S_{2,3} \cup S_{2,4} \cup S_{2,5}$	$S_{2,1}$	0.358	0.667	0.371	1.403	0.360	1.530	0.634	0.114	2.825
		$S_{2,2}$	0.198	1.102	0.697	0.000	0.000	1.734	0.110	0.595	2.632
		$S_{2,3}$	0.532	0.422	0.323	1.676	1.009	0.000	0.180	0.020	2.458
		$S_{2,4}$	0.528	0.000	0.430	0.876	0.579	0.461	0.435	0.040	0.694
		$S_{2,5}$	0.659	1.408	0.516	0.000	0.382	0.000	0.173	0.064	1.570

Table 5: Comparing results for the Swallow data set (R_1, R_2).

Items	Proposed method	Decision tree	Hyperplane support vector
$AR(R_1, R_2)$	(1, 1)	(0.97, 1)	(0.97, 1)
$SR(R_1, R_2)$	(1, 0.97)	(0.97, 1)	(0.97, 1)
CR	1	0.3	0.1

The accuracy rates of (R_1, R_2, R_3) in the proposed method are (1,1,1), as Model 1 has been solved. This finding indicates that none of objects in class 2 or class 3 are covered by $S_{1,1}$, none of objects in classes 1 or 3 are covered by $S_{2,1} \cup S_{2,2} \cup S_{2,3}$, and none of the objects in classes 1 or 2 are covered by $S_{3,1} \cup S_{3,2}$. The support rate of (R_1, R_2, R_3) in the proposed method is (1,0.98,0.98), indicating that all objects in class 1 are covered by $S_{1,1}$, 98% of the objects in class 2 are covered by $S_{2,1}$, $S_{2,2}$, and $S_{2,3}$, and 98% of the objects in class 3 are covered by $S_{3,1}$ and $S_{3,2}$. The compactness rate of rules R_1 , R_2 , and R_3 is computed as $CR(R_1, R_2, R_3) = 3/3 = 1$. Finally, we determine the following.

- (i) Although all three methods perform very well in the rates of accuracy and support, the proposed method has the best performance for the accuracy of classes 2 and 3 (i.e., R_2 and R_3).
- (ii) The proposed method has the best compactness rate.

4.2. Swallow Dataset

The European barn swallow (*Hirundo rustica*) dataset was obtained by trapping individual swallows in Stirlingshire, Scotland, between May and July 1997. This dataset contains 69 swallows. Each object is described by eight attributes, and it belongs to one of two classes (i.e., the birds are classified by the gender of individual birds).

Here, we also used Model 1 to induce the classification rules. Table 4 lists the optimal solutions (i.e., centroid and radius) for both rules R_1 and R_2 .

The result of the decision tree method, which is referred to in Li and Chen [3], is listed in Table 5, where $AR(R_1, R_2) = (0.97, 1)$, $SR(R_1, R_2) = (0.97, 1)$, and $CR = 0.3$.

The result of the hyperplane method, referred to in Chang and Lin [35], is also listed in Table 5, where $AR(R_1, R_2) = (0.97, 1)$, $SR(R_1, R_2) = (0.97, 1)$, and $CR = 0.1$.

We compared the three methods in Table 5 to show that the proposed method can induce rules with better or equivalent values of AR and SR . In fact, the proposed method also has the best compactness rate.

4.3. HSV Dataset

The HSV dataset contains 122 patients classified into four classes, with each patient having 11 preoperating attributes. To maximize the support rate with respect to the proposed method (i.e., Model 1), the proposed method generated seven hyperspheres and three unions of hyperspheres. The centroids and radii of the hyperspheres are reported in Table 6, and a comparison with other methods is reported in Table 7.

Using the decision tree method in the HSV dataset generates 24 rules. In addition, the hyperplane method deduces 45 hyperplanes for the HSV dataset. Table 7 also shows that the proposed method can find rules with the highest rates (i.e., AR , SR , and CR) compared with the other two methods.

4.4. Limitation of the Proposed Method

The hypersphere models are solved by one of the most powerful mixed-integer program software CPLEX [32] running in a PC. Based on optimization technique, the results of the numerical examples illustrate that the usefulness of the proposed method is better than that of the current methods, including the decision tree method and the hyperplane support vector method. As the solving time of the hypersphere model, which is linearized, mainly depends on the number of binary variables and constraints, solving the reformulated hypersphere model from the proposed algorithm takes about one minute for each dataset (i.e., in Sections 4.1 and 4.3), in which using eight piecewise line segments linearizes the nonlinear nonconvex term (i.e., $L(h_{i,k,j}^2)$) of Model 1.

The computing time for solving a linearized hypersphere program grows rapidly as the numbers of binary variables and constraints increase. Also, the computing time of the proposed method is slower than that of the decision tree method and hyperplane method, especially for large datasets or a great number of piecewise line segments. In the further study, utilizing a mainframe-version optimization software [36–38], integrating meta-heuristic algorithms, or using distributed computing techniques can enhance solving speed to conquer this problem.

5. Conclusions and Future Work

This study proposes a novel method for deducing classification rules, which can find the optimal solution based on a hypersphere domain. The optimization technique for finding classification rules is approached to optimal. Results of the numerical examples illustrate that the usefulness of the proposed method is better than that of the current methods, including the decision tree method and the hyperplane method. The proposed method is guaranteed to find an optimal rule, but the computational complexity grows rapidly by increasing the problem size. More investigation and research are required to enhance further

Table 6: Centroid points for the HSV data by the proposed method.

Rule numbers	Union of spheres	$S_{k,l}$	$h_{t,k,1}$	$h_{t,k,2}$	$h_{t,k,3}$	$h_{t,k,4}$	$h_{t,k,5}$	$h_{t,k,6}$	$h_{t,k,7}$	$h_{t,k,8}$	$h_{t,k,9}$	$h_{t,k,10}$	$h_{t,k,11}$	$r_{t,k}$
R_1	$S_{1,1} \cup S_{1,2} \dots \cup S_{1,8}$	$S_{1,1}$	0.504	0.528	0.366	0.288	0.850	0.848	0.351	-0.282	0.405	0.234	0.313	1.576
		$S_{1,2}$	0.469	0.395	0.312	0.133	-0.274	-0.309	0.134	0.590	0.813	1.000	-0.847	3.044
		$S_{1,3}$	0.287	0.867	0.458	0.189	0.977	0.267	0.017	-0.846	0.462	-1.000	1.000	4.249
		$S_{1,4}$	0.586	-0.400	0.605	0.467	-0.167	-0.631	0.268	1.000	0.526	1.000	0.120	3.365
		$S_{1,5}$	0.296	-0.511	-1.000	0.536	-0.263	0.552	0.954	0.635	-0.117	-1.000	0.600	4.608
		$S_{1,6}$	0.775	0.678	-0.092	0.340	0.366	0.455	0.451	1.000	0.422	0.295	0.470	1.560
		$S_{1,7}$	0.525	0.296	0.194	0.124	-0.945	-0.467	0.127	1.000	0.097	-0.897	0.654	4.113
		$S_{1,8}$	0.000	0.140	0.109	0.250	0.193	0.168	0.063	0.142	0.181	0.119	0.082	0.089
R_2	$S_{2,1} \cup S_{2,2} \cup S_{2,3} \cup S_{2,4}$	$S_{2,1}$	0.513	0.056	0.143	0.078	0.533	0.439	0.180	0.167	-0.284	-0.464	0.749	1.450
		$S_{2,2}$	0.533	0.381	0.139	0.568	-0.373	-0.005	0.080	0.831	0.293	0.215	0.369	1.293
		$S_{2,3}$	0.000	-0.450	0.709	0.553	0.896	1.000	0.157	-0.447	0.068	-0.580	1.000	3.579
		$S_{2,4}$	0.000	0.862	0.543	-0.066	0.389	0.451	0.106	0.394	-0.004	0.042	-0.014	0.409
R_3	$S_{3,1} \cup S_{3,2} \cup S_{3,3}$	$S_{3,1}$	0.624	0.507	-1.000	0.147	0.840	1.000	0.827	-0.585	0.199	0.831	-0.380	4.277
		$S_{3,2}$	0.534	0.483	-0.088	0.365	-0.364	0.270	0.437	0.785	0.786	0.682	-1.000	3.003
		$S_{3,3}$	0.000	0.210	0.630	0.750	-1.000	0.244	0.388	1.000	0.474	-0.604	0.475	3.162
R_4	$S_{4,1} \cup S_{4,2} \cup S_{4,3} \cup S_{4,4} \cup S_{4,5}$	$S_{4,1}$	0.551	0.374	0.637	0.256	0.865	0.944	0.315	-0.831	-0.485	-0.676	0.979	4.330
		$S_{4,2}$	0.717	0.254	0.548	-0.718	0.118	0.730	-0.547	0.498	-0.580	-0.464	0.821	3.287
		$S_{4,3}$	0.527	1.000	0.533	0.089	-0.152	-1.000	0.366	0.046	0.547	0.947	-0.436	2.943
		$S_{4,4}$	0.489	0.625	-0.209	0.522	0.644	0.582	0.615	1.000	0.306	-0.401	0.931	2.059
		$S_{4,5}$	-0.011	0.491	-0.409	0.155	-0.766	-0.227	-0.107	1.000	0.303	-0.680	-0.185	2.843

Table 7: Comparing results for the HSV data set (R_1, R_2, R_3, R_4) .

Items	Proposed method	Decision tree	Hyperplane support vector
$AR(R_1, R_2, R_3, R_4)$	(1,1,1,1)	(0.93,0.81,0.7,0.71)	(0.9,1,1,0.9)
$SR(R_1, R_2, R_3, R_4)$	(0.99,1,1,1)	(0.93,0.72,0.78,0.71)	(0.9,0.72,0.67,0.69)
CR	0.4	0.17	0.09

the computational efficiency of globally solving large-scale classification problems, such as running mainframe-version optimization software, integrating meta-heuristic algorithms, or using distributed computing techniques.

Acknowledgments

The authors wish to thank the editor and the anonymous referees for providing insightful comments and suggestions, which have helped them improving the quality of the paper. This work was supported by the National Science Council of Taiwan under Grants NSC 100-2811-E-009-040-, NSC 99-2221-E-030-005-, and NSC 100-2221-E-030-009-.

References

- [1] M. J. Beynon and K. L. Buchanan, "An illustration of variable precision rough set theory: the gender classification of the European barn swallow (*Hirundo rustica*)," *Bulletin of Mathematical Biology*, vol. 65, no. 5, pp. 835–858, 2003.
- [2] H. L. Li and C. J. Fu, "A linear programming approach for identifying a consensus sequence on DNA sequences," *Bioinformatics*, vol. 21, no. 9, pp. 1838–1845, 2005.
- [3] H. L. Li and M. H. Chen, "Induction of multiple criteria optimal classification rules for biological and medical data," *Computers in Biology and Medicine*, vol. 38, no. 1, pp. 42–52, 2008.
- [4] C. W. Chu, G. S. Liang, and C. T. Liao, "Controlling inventory by combining ABC analysis and fuzzy classification," *Computers and Industrial Engineering*, vol. 55, no. 4, pp. 841–851, 2008.
- [5] J.-X. Chen, "Peer-estimation for multiple criteria ABC inventory classification," *Computers & Operations Research*, vol. 38, no. 12, pp. 1784–1791, 2011.
- [6] E. B. Hunt, J. Marin, and P. J. Stone, *Experiments in Induction*, Academic Press, New York, NY, USA, 1966.
- [7] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth Statistics/Probability Series, Wadsworth Advanced Books and Software, Belmont, Calif, USA, 1984.
- [8] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [9] J. R. Quinlan, "Simplifying decision trees," *International Journal of Man-Machine Studies*, vol. 27, no. 3, pp. 221–234, 1987.
- [10] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufman, San Mateo, Calif, USA, 1993.
- [11] H. Kim and W. Y. Loh, "Classification Trees with Unbiased Multiway Splits," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 589–604, 2001.
- [12] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [13] R. M. Rifkin, *Everything old is new again: a fresh look at historical approaches in machine learning*, Ph.D. thesis, Massachusetts Institute of Technology, Ann Arbor, MI, USA, 2002.
- [14] S. Katagiri and S. Abe, "Incremental training of support vector machines using hyperspheres," *Pattern Recognition Letters*, vol. 27, no. 13, pp. 1495–1507, 2006.
- [15] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

- [16] C. H. Chu and D. Widjaja, "Neural network system for forecasting method selection," *Decision Support Systems*, vol. 12, no. 1, pp. 13–24, 1994.
- [17] C. Giraud-Carrier, R. Vilalta, and P. Brazdil, "Guest editorial: introduction to the special issue on meta-learning," *Machine Learning*, vol. 54, no. 3, pp. 187–193, 2004.
- [18] D. Pomorski and P. B. Perche, "Inductive learning of decision trees: application to fault isolation of an induction motor," *Engineering Applications of Artificial Intelligence*, vol. 14, no. 2, pp. 155–166, 2001.
- [19] H. Zhang, C. N. Yu, B. Singer, and M. Xiong, "Recursive partitioning for tumor classification with gene expression microarray data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 12, pp. 6730–6735, 2001.
- [20] S. Aja-Fernández, R. De Luis-García, M. Á. Martín-Fernández, and C. Alberola-López, "A computational TW3 classifier for skeletal maturity assessment. A Computing with Words approach," *Journal of Biomedical Informatics*, vol. 37, no. 2, pp. 99–107, 2004.
- [21] P. Geurts, M. Fillet, D. de Seny et al., "Proteomic mass spectra classification using decision tree based ensemble methods," *Bioinformatics*, vol. 21, no. 14, pp. 3138–3145, 2005.
- [22] C. Olaru and L. Wehenkel, "A complete fuzzy decision tree technique," *Fuzzy Sets and Systems*, vol. 138, no. 2, pp. 221–254, 2003.
- [23] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sensing of Environment*, vol. 86, no. 4, pp. 554–565, 2003.
- [24] Z. Chen, J. Li, and L. Wei, "A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue," *Artificial Intelligence in Medicine*, vol. 41, no. 2, pp. 161–175, 2007.
- [25] H. Nunez, C. Angulo, and A. Catala, "Rule-extraction from support vector machines," in *Proceedings of the European Symposium on Artificial Neural Networks*, pp. 107–112, 2002.
- [26] Y. M. Lin, X. Wang, W. W. Y. Ng, Q. Chang, D. S. Yeung, and X. L. Wang, "Sphere classification for ambiguous data," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 2571–2574, August 2006.
- [27] J. Wang, P. Neskovic, and L. N. Cooper, "Bayes classification based on minimum bounding spheres," *Neurocomputing*, vol. 70, no. 4–6, pp. 801–808, 2007.
- [28] L. Gu and H. Z. Wu, "A kernel-based fuzzy greedy multiple hyperspheres covering algorithm for pattern classification," *Neurocomputing*, vol. 72, no. 1–3, pp. 313–320, 2008.
- [29] M. Hifi and R. M'Hallah, "A literature review on circle and sphere packing problems: models and methodologies," *Advances in Operations Research*, vol. 2009, Article ID 150624, 22 pages, 2009.
- [30] M. Zhu, Y. Wang, S. Chen, and X. Liu, "Sphere-structured support vector machines for multi-class pattern recognition," in *Proceedings of the 9th International Conference (RSFDGrC '03)*, vol. 2639 of *Lecture Notes in Computer Science*, pp. 589–593, May 2003.
- [31] E. M. L. Beale and J. J. H. Forrest, "Global optimization using special ordered sets," *Mathematical Programming*, vol. 10, no. 1, pp. 52–69, 1976.
- [32] IBM/ILOG, "Cplex 12.0 reference manual," 2010, <http://www.ilog.com/products/cplex/>.
- [33] D. C. Dunn, W. E. G. Thomas, and J. O. Hunter, "An evaluation of highly selective vagotomy in the treatment of chronic duodenal ulcer," *Surgery Gynecology and Obstetrics*, vol. 150, no. 6, pp. 845–849, 1980.
- [34] K. Slowinski, "Rough classification of HSV patients," in *Intelligent Decision Support-Handbook of Applications and Advances of the Rough Sets Theory*, R. Slowinski, Ed., pp. 77–94, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1992.
- [35] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," 2010, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html>.
- [36] C. Than, R. Sugino, H. Innan, and L. Nakhleh, "Efficient inference of bacterial strain trees from genome-scale multilocus data," *Bioinformatics*, vol. 24, no. 13, pp. i123–i131, 2008.
- [37] M. Rockville, "Large scale computing and storage requirements for biological and environmental research," Tech. Rep., Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, Calif, USA, 2009.
- [38] X. Xie, X. Fang, S. Hu, and D. Wu, "Evolution of supercomputers," *Frontiers of Computer Science in China*, vol. 4, no. 4, pp. 428–436, 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

