Pattern Anal Applic (2012) 15:237–247
DOI 10.1007/s10044-011-0208-z

THEORETICAL ADVANCES

An extension to fuzzy support vector data description (FSVDD*)

Y. Forghani · H. Sadoghi Yazdi · S. Effati

Received: 7 April 2010 / Accepted: 5 March 2011 / Published online: 22 March 2011

Abstract The well-known support vector data description (SVDD) is based on precise description of precise data. When we know the features of training samples precisely and we are uncertain about their class labels, the fuzzy SVDD can be used to obtain the data description. But if the features of training samples are fuzzy numbers, the fuzzy SVDD cannot be utilized. In this paper, we extend the fuzzy SVDD for the description of such training samples and then apply our proposed method, called FSVDD*, to real data. The experimental results show the ability of the proposed method in Taiwanese tea evaluation.

Keywords Fuzzy SVDD · Data description · Fuzzy numbers · Distance · Defuzzification

1 Introduction

Pursuing of effective methods for real data analysis is considered with researchers recently. Progress in fuzzy mathematics encourages engineering and other scientists in application fields for presentation of real models such as fuzzy image filtering [1], fuzzy clustering [2, 3], fuzzy multivariable nonlinear regression analysis [2] and fuzzy classification [3, 4] (Classification is among the most important problem tasks in the realm of data analysis, data mining and machine learning and has many applications in industry, including, e.g., oil spill detection [5], intrusion detection in computer networks [6], breast cancer detection [7], fingerprint identification [8], text document classification [9, 10], handwritten Tamil character recognition [11], Epo doping control [12], human identification [13, 14] and signature verification [15]).

In this paper, we want to introduce a one-class classifier based on support vector data description (SVDD) [16, 17] and its extended version, the fuzzy SVDD (FSVDD) [18], suited for working with real data which are usually uncertain. The FSVDD is a suitable approach in the field of pattern recognition for one-class classification or data description. In the FSVDD, one can assign different degree of importance to each training samples. The FSVDD includes quadratic programming with quadratic constraints. Thus, to solve the FSVDD for samples whose features are fuzzy numbers, a fuzzy quadratic program with fuzzy quadratic constraint must be solved. Specifically, nonlinear programming with crisp parameters has been widely used in solving real problems and several efforts reported in literature developing efficient algorithms for solving these types of problems [19]. Liu [20] solved a special form of fuzzy quadratic program by formulating a pair of two-level mathematical programs to calculate the upper bound and lower bound of the objective function of the fuzzy quadratic program, but using the Liu's method leads to obtain a non-convex problem which obtaining its global optimal solution is very time-consuming. Thus, we use a defuzzification method to solve the quadratic fuzzy programming. The defuzzification of fuzzy parameters of the fuzzy

Y. Forghani (✉) · H. Sadoghi Yazdi
Computer Department, Ferdowsi University of Mashhad, Mashhad, Iran
e-mail: yahyafor2000@yahoo.com

H. Sadoghi Yazdi
e-mail: h-sadoghi@um.ac.ir

S. Effati
Applied Mathematics Department, Ferdowsi University of Mashhad, Mashhad, Iran
e-mail: s-effati@um.ac.ir

quadratic program leads to obtain a crisp quadratic program which gets a crisp description of fuzzy data.

Organization of this paper is as follows: Section 2 will deal with some preliminaries and in Sect. 3 our novel method (FSVDD*) will be explained. Section 4 shows experimental results of the presented work. Finally, Sect. 5 concludes the paper.

2 Preliminaries

2.1 FSVDD

In one-class classification, if the class region is approximated by some method then if we test whether test samples are outside the region, the outliers can be detected. The approximation of the genuine class region is called the domain description. Tax and Duin extended the support vector method to domain description (SVDD). In this subsection, we discuss an extended version of their method, namely the FSVDD [18].

Let x_i ($i = 1, \ldots, n$) be p-dimensional training samples belonging to one class. We consider approximating the class region by the minimum hypersphere with center $e = (e_1, e_2, \ldots, e_p)^{\mathrm{T}}$ and radius R in high dimensional feature space (HDS), excluding the outliers. Then the problem is

$$\min_{R,e,\xi} R^2 + C\sum_{i=1}^{n} w_i\xi_i$$
$$\text{subject to}\begin{cases} \|g(x_i) - e\|^2 \le R^2 + \xi_i, & i = 1,\ldots,n, \\ \xi_i \ge 0, & i = 1,\ldots,n, \end{cases} \quad (1)$$

where $g(x)$ is the mapping function that maps x into a high dimension space (HDS), $\xi = (\xi_1, \ldots, \xi_n)^{\mathrm{T}}$ and ξ_i is the slack variable of ith training sample and w_i is its weight or its importance and C is a constant which determines the trade-off between the hypersphere volume and outliers. The Lagrangian dual form of (1) is as follows:

$$\max_{\delta} \sum_{i=1}^{n} \delta_i K(x_i, x_i) - \sum_{i=1}^{n}\sum_{j=1}^{n} \delta_i\delta_j K(x_i, x_j)$$
$$\text{subject to}\begin{cases} \sum_{i=1}^{n} \delta_i = 1, \\ 0 \le \delta_i \le Cw_i, \quad i = 1,\ldots,n, \end{cases} \quad (2)$$

which is a convex quadratic program and its global optimal solution can be obtained easily. After solving the program (2), it can be said that the unknown datum x is inside the hypersphere if $\|g(x) - e\|^2 \le R^2$ or equivalently if

$$K(x,x) - 2\sum_{i=1}^{n} \delta_i K(x, x_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} \delta_i\delta_j K(x_i, x_j) \le R^2, \quad (3)$$

where

$$R^2 = K(x_i, x_i) - 2\sum_{j=1}^{n} \delta_j K(x_i, x_j) + \sum_{j=1}^{n}\sum_{k=1}^{n} \delta_j\delta_k K(x_j, x_k) \quad \text{for } 0 < \delta_i < Cw_i. \quad (4)$$

See Appendix for more information.

2.2 Some fuzzy concepts

Definition 2.2.1 Let $\mathbb{R}$ denote the set of all real numbers. A fuzzy number is a mapping $\bar{x} : \mathbb{R} \to [0, 1]$ with the following properties:

- $\tilde{x}$ is a normal fuzzy set, i.e., the core of $\tilde{x} = C(\tilde{x}) = \{x \in \mathbb{R} : \mu_{\tilde{x}}\{x\} = 1\}$ is not empty.
- $\mu_{\tilde{x}}(\cdot)$ is upper semi-continuous.
- $\tilde{x}$ is a convex fuzzy set, i.e., $\mu_{\tilde{x}}(\lambda x + (1-\lambda)y) \ge \min\{\mu_{\tilde{x}}(x), \mu_{\tilde{x}}(y)\}$, for all $x, y \in \mathbb{R}, \lambda \in [0, 1]$.
- The support of $\tilde{x}$, $S(\tilde{x}) = \{x \in \mathbb{R} : \mu_{\tilde{x}}(x) > 0\}$ and its closure $cl(S(\tilde{x}))$ is compact.

Definition 2.2.2 The LR-type fuzzy number is a special type of representation for fuzzy number. It is defined by two functions L (and R) which map $\mathbb{R}^+ \to [0, 1]$ and are decreasing shape functions and $L(0) = 1$, $L(1) = 0$; $\forall x < 1$:$L(x) > 0$ and $\forall x > 0$:$L(x) < 1$. A fuzzy number $\tilde{x}$ is of LR type if there exist reference functions L (for left), R (for right) and scalars Δ, $\nabla > 0$, with

$$\mu_{\tilde{x}}(x) = \begin{cases} L\left(\frac{m-x}{\Delta}\right) & x \le m, \\ R\left(\frac{x-m}{\nabla}\right) & x \ge m. \end{cases} \quad (5)$$

Here, m, called the mean value of $\tilde{x}$, is a real member and Δ and ∇ are called the left and right spreads, respectively. Here $\mu_{\tilde{x}}(x)$ is membership function of fuzzy number $\tilde{x}$, denoted by $(m, \Delta, \nabla)_{\mathrm{LR}}$ (Fig. 1).

Definition 2.2.3 Let each feature of the training sample $\tilde{x}$ be a fuzzy number with probably different width of uncertainty (Fig. 2). Such training sample at α-cut can be shown by a hyper-rectangle (HR). We name this hyper-rectangle

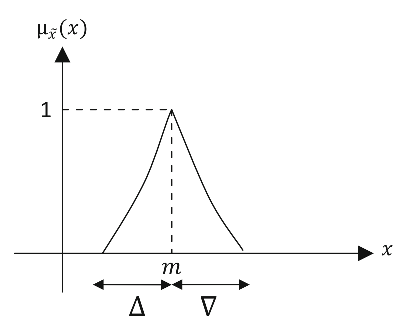

Fig. 1 An LR-type fuzzy number

as α-cut hyper-rectangle or α-cut HR. The width of each dimension of α-cut HR of this training sample shows the width of uncertainty of one of its feature at α-cut (see Fig. 3). In the other words, if the membership value of this fuzzy training sample is ignored, it can be shown by the set $\{(x_1, \ldots, x_p) | (x_k)_\alpha^{\mathrm{L}} \le x_k \le (x_k)_\alpha^{\mathrm{U}}, \quad k = 1, \ldots, p\}$, which is a hyper-rectangle, called α-cut HR.

2.3 Distance between two fuzzy numbers

Let us consider distances for fuzzy numbers. Some of these distances were proposed by [21–23]. A method of the fuzzy data preprocessing is based on these distances. Indeed, these metrics can be used for defuzzification of the distance between two fuzzy numbers.

Definition 2.3.1 (*The Yang Distance*) The Yang distance [23] for two LR-type fuzzy numbers $\tilde{x} = (m_{\tilde{x}}, \Delta_{\tilde{x}}, \nabla_{\tilde{x}})_{\mathrm{LR}}$ and $\tilde{y} = (m_{\tilde{y}}, \Delta_{\tilde{y}}, \nabla_{\tilde{y}})_{\mathrm{LR}}$ is as follows:

$$d^2_{\mathrm{Yang}}(\tilde{x}, \tilde{y}) = (m_{\tilde{x}} - m_{\tilde{y}})^2 + ((m_{\tilde{x}} - l\Delta_{\tilde{x}}) - (m_{\tilde{y}} - l\Delta_{\tilde{y}}))^2 + ((m_{\tilde{x}} + r\nabla_{\tilde{x}}) - (m_{\tilde{y}} + r\nabla_{\tilde{y}}))^2 \tag{6}$$

where $l = \int_0^1 L^{-1}(w)\,\mathrm{d}w$ and $r = \int_0^1 R^{-1}(w)\,\mathrm{d}w$. The Yang distance for two triangular fuzzy numbers $\tilde{x} = (m_{\tilde{x}}, \Delta_{\tilde{x}}, \nabla_{\tilde{x}})_{\mathrm{T}}$ and$\tilde{y} = (m_{\tilde{y}}, \Delta_{\tilde{y}}, \nabla_{\tilde{y}})_{\mathrm{T}}$ becomes as follows:

$$d^2_{\mathrm{Yang}}(\tilde{x}, \tilde{y}) = (m_{\tilde{x}} - m_{\tilde{y}})^2 + \left((m_{\tilde{x}} - m_{\tilde{y}}) - \frac{1}{2}(\Delta_{\tilde{x}} - \Delta_{\tilde{y}})\right)^2 + \left((m_{\tilde{x}} - m_{\tilde{y}}) - \frac{1}{2}(\Delta_{\tilde{x}} - \nabla_{\tilde{y}})\right)^2 \tag{7}$$

Definition 2.3.2 (*The Hausdorff Distance*) For any two fuzzy numbers $\tilde{x}$ and $\tilde{y}$, the Hausdorff distance metric is defined by [21] as follows:

$$d^2_{\mathrm{Hausdorff}}(\tilde{x}, \tilde{y}) = \max\left\{\|\tilde{x}_\alpha^{\mathrm{L}} - \tilde{y}_\alpha^{\mathrm{L}}\|^2, \|\tilde{x}_\alpha^{\mathrm{U}} - \tilde{y}_\alpha^{\mathrm{U}}\|^2\right\}, \tag{8}$$

where $\tilde{x}_\alpha = [\tilde{x}_\alpha^{\mathrm{L}}, \tilde{x}_\alpha^{\mathrm{U}}] = \{x : \mu_{\tilde{x}}(x) \ge \alpha\}$ and $\alpha \in [0, 1]$.

Definition 2.3.3 (*The Hathaway Distance*) Let $\tilde{x} = (m_{\tilde{x}}, d_{\tilde{x}}, \Delta_{\tilde{x}}, \nabla_{\tilde{x}})_{\mathrm{T1}}$ and $\tilde{y} = (m_{\tilde{y}}, d_{\tilde{y}}, \Delta_{\tilde{y}}, \nabla_{\tilde{y}})_{\mathrm{T1}}$ be two trapezoidal fuzzy numbers. Then the Hathaway distance is defined by [22] as follows:

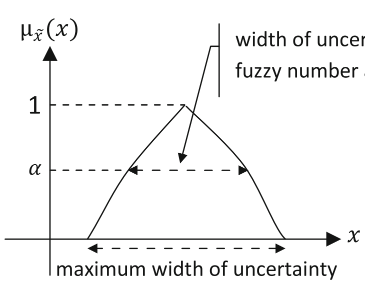

Fig. 2 Width of uncertainty

$$d^2_{\mathrm{Hathaway}}(\tilde{x}, \tilde{y}) = (m_{\tilde{x}} - m_{\tilde{y}})^2 + (d_{\tilde{x}} - d_{\tilde{y}})^2 + (\Delta_{\tilde{x}} - \Delta_{\tilde{y}})^2 + (\nabla_{\tilde{x}} - \nabla_{\tilde{y}})^2. \tag{9}$$

3 Our novel method

The well-known SVDD is based on precise description of precise data. When we know the features of a training sample precisely and we are uncertain about its class label, the FSVDD can be used for data description and low value is assigned to the weight of such training sample, but when some features of training samples are fuzzy numbers, the FSVDD cannot be utilized to obtain the data description. In our proposed method (FSVDD*), the description of such training samples can be obtained.

Let x_{ik} be kth feature of ith training sample $x_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$. We suppose that the features of training samples, x_{ik}, are approximately known and can be represented by LR-type fuzzy number $\tilde{x}_{ik}$. Thus, the formulation of our novel method can be stated as follows:

$$\min_{\tilde{R}, \tilde{e}, \tilde{\xi}} \tilde{R}^2 + C\sum_{i=1}^{n} w_i \tilde{\xi}_i$$
$$\text{subject to} \begin{cases} \|g(\tilde{x}_i) - \tilde{e}\|^2 \le \tilde{R}^2 + \tilde{\xi}_i, & i = 1, \ldots, n, \\ \tilde{\xi}_i \ge 0, & i = 1, \ldots, n, \end{cases} \tag{10}$$

where $\tilde{\xi} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_n)^{\mathrm{T}}$, $\tilde{e} = (\tilde{e}_1, \ldots, \tilde{e}_n)^{\mathrm{T}}$ and $\tilde{R}$ are fuzzy set. One approach to solve this program is to use a defuzzification method. That is, to solve the program (10) it suffices to convert it to a crisp program as follows:

$$\min_{D(\tilde{R}), D(\tilde{e}), D(\tilde{\xi})} D(\tilde{R})^2 + C\sum_{i=1}^{n} w_i D(\tilde{\xi}_i)$$
$$\text{subject to} \begin{cases} \|D(g(\tilde{x}_i)) - D(\tilde{e})\|^2 \le D(\tilde{R})^2 + D(\tilde{\xi}_i), & i = 1, \ldots, n, \\ D(\tilde{\xi}) \ge 0, & i = 1, \ldots, n, \end{cases} \tag{11}$$

where $D(\cdot)$ is a defuzzification function. The Lagrangian dual form of (11) is

$$\max_{\delta, \gamma} L(\tilde{R}, \tilde{e}, \tilde{\xi}, \delta, \gamma)$$
$$\text{subject to } \delta_i, \gamma_i \ge 0, \quad i = 1, \ldots, n, \tag{12}$$

where $\delta = (\delta_1, \ldots, \delta_n)^{\mathrm{T}}$, $\gamma = (\gamma_1, \ldots, \gamma_n)^{\mathrm{T}}$ and

$$L(\tilde{R}, \tilde{e}, \tilde{\xi}, \delta, \gamma) = D(\tilde{R})^2 + C\sum_{i=1}^{n} w_i D(\tilde{\xi}_i) - \sum_{i=1}^{n} \delta_i (D(\tilde{R})^2 + D(\tilde{\xi}_i) - D(g(\tilde{x}_i))^{\mathrm{T}} D(g(\tilde{x}_i)) + 2D(\tilde{e})^{\mathrm{T}} D(g(\tilde{x}_i)) - D(\tilde{e})^{\mathrm{T}} D(\tilde{e})) - \sum_{i=1}^{n} \gamma_i D(\tilde{\xi}_i). \tag{13}$$

For the optimal solution, the following conditions are satisfied

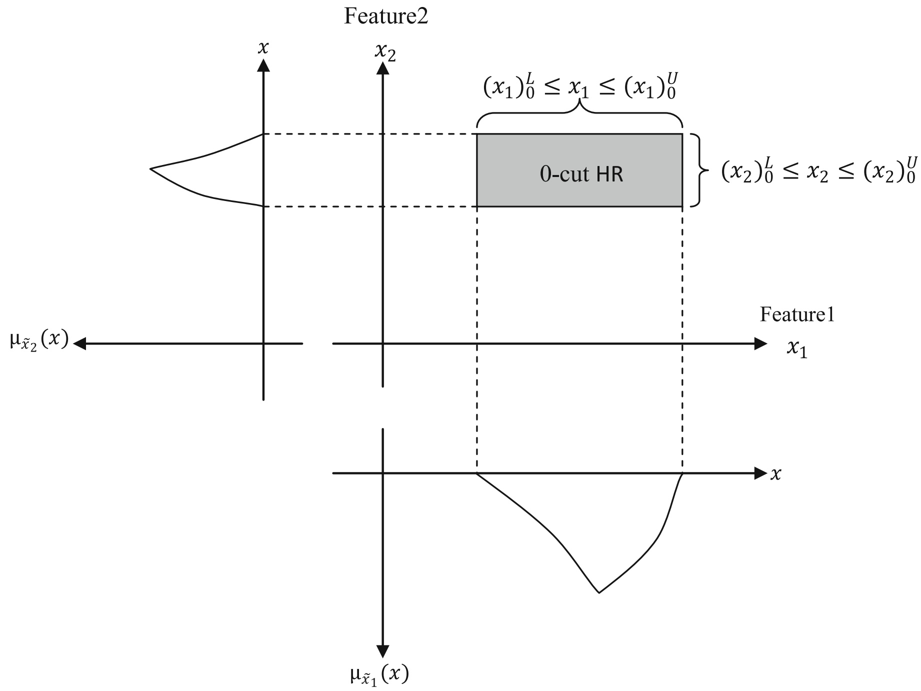

Fig. 3 0-cut HR of a two-dimensional fuzzy sample $\tilde{x} = (\tilde{x}_1, \tilde{x}_2)^{\mathrm{T}}$

$$\frac{\partial L}{\partial D(\tilde{R})} = 0 \rightarrow \sum_{i=1}^{n} \delta_i = 1, \tag{14}$$

$$\frac{\partial L}{\partial D(\tilde{e})} = 0 \rightarrow D(\tilde{e}) = \sum_{i=1}^{n} \delta_i D(g(\tilde{x}_i)), \tag{15}$$

$$\frac{\partial L}{\partial D(\tilde{\xi})} = 0 \rightarrow \delta_i = Cw_i - \gamma_i, \quad i = 1, \ldots, n, \tag{16}$$

$$\delta_i(\|D(g(\tilde{x}_i)) - D(\tilde{e})\|^2 - D(\tilde{R})^2 - D(\tilde{\xi}_i)) = 0, \quad i = 1, \ldots, n, \tag{17}$$

$$\gamma_i D(\tilde{\xi}_i) = 0, \quad i = 1, \ldots, n. \tag{18}$$

Using the above conditions, $L(\tilde{R}, \tilde{e}, \tilde{\xi}, \delta, \gamma)$ is transformed to

$$\sum_{i=1}^{n} \delta_i K_r(x_i, x_i) - \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j K_r(x_i, \tilde{x}_j), \tag{19}$$

where

$$K_r(\tilde{x}_i, \tilde{x}_j) = D(g(\tilde{x}_i))^{\mathrm{T}} D(g(\tilde{x}_i)) = D^*(g(\tilde{x}_i)^{\mathrm{T}} g(\tilde{x}_j)) = D^*(K(\tilde{x}_i, \tilde{x}_j)), \tag{20}$$

and $D^*(\cdot)$ is a defuzzification function. Since $\delta_i \ge 0$ and from (16) we have $0 \le \delta_i \le Cw_i$. Thus, the Lagrangian dual form of (11) can be stated as follows:

$$\max_\delta \sum_{i=1}^{n} \delta_i K_r(\tilde{x}_i, \tilde{x}_i) - \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j K_r(\tilde{x}_i, \tilde{x}_j)$$
$$\text{subject to} \begin{cases} \sum_{i=1}^{n} \delta_i = 1, \\ 0 \le \delta_i \le Cw_i, \quad i = 1, \ldots, n. \end{cases} \tag{21}$$

For example and without loss of generality let to use Gaussian kernel function, i.e., $K(\tilde{x}, \tilde{y}) = e^{\frac{-\|\tilde{x}-\tilde{y}\|^2}{2\sigma^2}}$. Thus, $K_r(\tilde{x}, \tilde{y}) = D^*(K(\tilde{x}_i, \tilde{x}_j)) = D^*\left(e^{\frac{-\|\tilde{x}-\tilde{y}\|^2}{2\sigma^2}}\right) = e^{\frac{-D^{**}(\|\tilde{x}-\tilde{y}\|^2)}{2\sigma^2}}$ and the program (21) can be stated as follows:

$$\min_\delta \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j e^{\frac{-D^{**}(\|\tilde{x}_i - \tilde{x}_j\|^2)}{2\sigma^2}}$$
$$\text{subject to} \begin{cases} \sum_{i=1}^{n} \delta_i = 1, \\ 0 \le \delta_i \le Cw_i, \quad i = 1, \ldots, n. \end{cases} \tag{22}$$

Now, since $\tilde{x}_{ik} = (m_{\tilde{x}ik}, \Delta_{\tilde{x}ik}, \nabla_{\tilde{x}ik})_{\mathrm{LR}}$ was considered to be LR-type fuzzy number and without the loss of generality, the following defuzzification function can be used

$$D^{**}(\|\tilde{x}_i - \tilde{x}_j\|^2) = d^2_{\mathrm{LR}}(\tilde{x}_i, \tilde{x}_j), \tag{23}$$

where $d^2_{\mathrm{LR}}(\cdot, \cdot)$ is a crisp distance between two LR-type fuzzy vectors such as the Yang distance. If the Yang distance is used, the program (22) is restated as follows:

$$\min_\delta \sum_{i=1}^{n} \sum_{j=1}^{n}$$
$$\times \delta_i \delta_j e^{\frac{-((m_{x_i} - m_{x_j})^2 + ((m_{x_i} - l\Delta_{x_i}) - (m_{x_j} - l\Delta_{x_j}))^2 + ((m_{x_i} - r\nabla_{x_i}) - (m_{x_j} - r\nabla_{x_j}))^2)}{2\sigma^2}}$$
$$\text{subject to} \begin{cases} \sum_{i=1}^{n} \delta_i = 1 \\ 0 \le \delta_i \le Cw_i, \quad i, \ldots, n, \end{cases} \tag{24}$$

where $m_{\tilde{x}_i} = (m_{\tilde{x}_{i1}}, \ldots, m_{\tilde{x}_{ip}})^{\mathrm{T}}, \Delta_{\tilde{x}_i} = (\Delta_{\tilde{x}_{i1}}, \ldots, \Delta_{\tilde{x}_{ip}})^{\mathrm{T}}$ and $\nabla_{\tilde{x}_i} = (\nabla_{\tilde{x}_{i1}}, \ldots, \nabla_{\tilde{x}_{ip}})^{\mathrm{T}}$. Now, we have a convex quadratic program with crisp parameters. That is, its local optimal solution is global optimal solution, too. Therefore, its global optimal solution can be obtained easily and there exists standard algorithm to solve it. From (15), we have $\|D(g(\tilde{x}_i)) - D(\tilde{e})\|^2 = K_r(\tilde{x}_i, \tilde{x}_i) - 2\sum_{j=1}^{n} \delta_j K_r(\tilde{x}_i, \tilde{x}_j) + \sum_{j=1}^{n}\sum_{k=1}^{n} \delta_i\delta_k K_r(\tilde{x}_j, \tilde{x}_k)$, and from (17), if $\delta_i > 0$, $K_r(\tilde{x}_i,\tilde{x}_i) - 2\sum_{j=1}^{n} \delta_j K_r(\tilde{x}_i,\tilde{x}_j) + \sum_{j=1}^{n}\sum_{k=1}^{n} \delta_j\delta_k K_r(\tilde{x}_j,\tilde{x}_k) = D(\tilde{R})^2 + D(\tilde{\xi}_i)$. From (16), if $\delta_i < Cw_i$ then $\gamma_i > 0$ and from (18) we have $\mathrm{Rank}(\tilde{\xi}_i) = 0$.Therefore, if $0 < \delta_i < Cw_i$,

$$R^2 = K_r(\tilde{x}_i, \tilde{x}_i) - 2\sum_{j=1}^{n} \delta_j K_r(\tilde{x}_i, \tilde{x}_j) + \sum_{j=1}^{n}\sum_{k=1}^{n} \delta_j\delta_k K_r(\tilde{x}_j, \tilde{x}_k). \tag{25}$$

Finally, the unknown datum $\tilde{x}$ is inside the hypersphere if $\|D(g(\tilde{x})) - D(\tilde{e})\|^2 - D(\tilde{R})^2 \leq 0$ or equivalently if

$$f(\tilde{x}; \delta, \tilde{R}) = K_r(\tilde{x}, \tilde{x}) - 2\sum_{i=1}^{n} \delta_i K_r(\tilde{x}, \tilde{x}_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} \delta_i\delta_j K_r(\tilde{x}_i, \tilde{x}_j) - D(\tilde{R})^2. \tag{26}$$

4 Experimental results

4.1 Numerical examples

In this section, the proposed algorithm is utilized using different training samples. Here, for ease of evaluation, two-dimensional data are used and each feature of training samples is considered to be a symmetric triangular fuzzy number. The cores of the training samples are selected from a banana distribution with variance of 7. Moreover, the Gaussian kernel function and the Yang distance are used to obtain the description of fuzzy data.

In the first experiment, 10 fuzzy training samples were used (see Fig. 4). The 0-cut HR of these training samples has been shown by rectangle and their cores by small stars. The thick curve shows the data description obtained using our proposed method. Then, we changed the core of one of the fuzzy training samples shown by dashed rectangle in Figs. 5 and 6 from its center to its corner and obtained the objective data description (shown by thin curve). As it can be seen, the proposed method is sensitive to the cores of fuzzy training samples.

In the next experiment, we used the same fuzzy training samples, but this time we changed the maximum width of uncertainty of a feature of one of the training samples shown by dashed rectangle (Figs. 7, 8, 9, 10). As it can be seen, the proposed method is also sensitive to the width of uncertainty of the features of the training samples, but increasing left or right spread parameter of a fuzzy feature of a training sample has just the same effect on the description of data, which is a drawback.

Our experiments show that the Hathaway distance also has the explained drawback, but the Hausdorff distance does not have. However, when we use the Hausdorff distance and changed the core of a training samples from its center to its corner as it explained earlier, the description of data does not change, which is also a drawback.

Meanwhile, the drawn curves for the description of data have been obtained using the crisp test data. In the other word, the crisp test samples which lie in the drawn curves are accepted as a member of the genuine class. To determine the classification of a fuzzy test sample, Eq. 26 must be used.

4.2 Application to one-class classification of Taiwanese Grade-one-tea

In this section, we use the proposed method for one-class classification of Taiwanese Grade-one-tea and use Taiwanese tea dataset [24]. To do so, we use the Yang distance for defuzzification.

Tea has become a more important agricultural product in Taiwan and there are currently 20,000 h of tea farms with an annual production of 21,000 tons. The types of tea produced in Taiwan include green tea, Paochong, Oolong and black tea. In recent years, the majority of teas produced in Taiwan have been of the Paochong and Oolong varieties. Black tea and green tea are relatively minor types in comparison. Because the tea varieties and

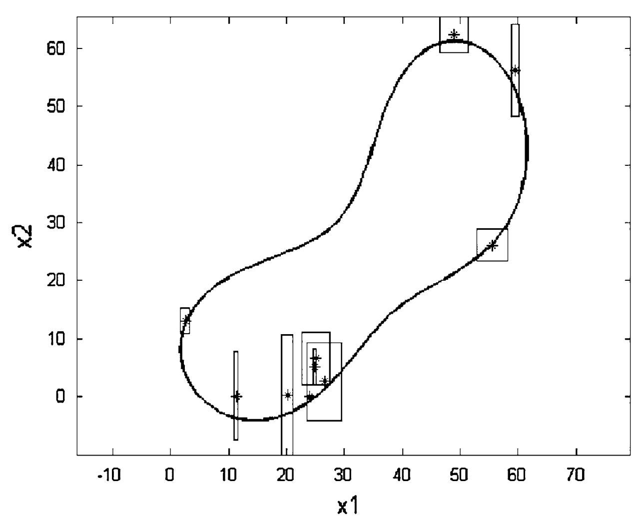

Fig. 4 The description of fuzzy training samples using the proposed method (the *rectangles* show 0-cut HR and the *small stars* show the cores of the training samples)

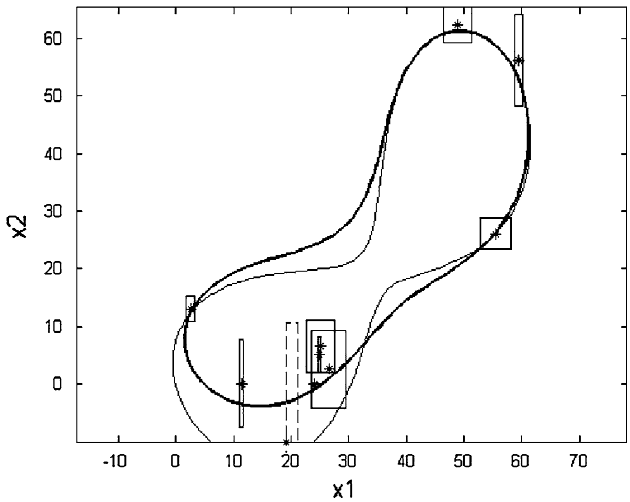

Fig. 5 *Thick curve* The description of data before changing the core of the fuzzy training sample (shown by *dashed rectangle*) from its center to its corner. *Thin curve* The description of data after changing the core of the fuzzy training sample

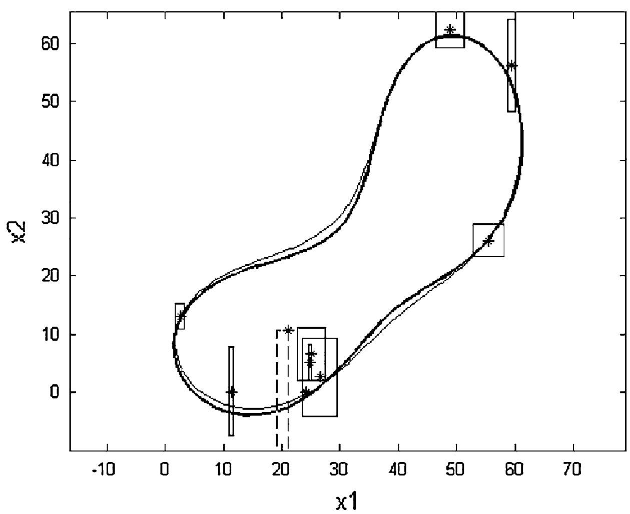

Fig. 6 *Thick curve* The description of data before changing the core of the fuzzy training sample (shown by *dashed rectangle*) from its center to its corner. *Thin curve* The description of data after changing the core of the fuzzy training sample

prices are numerous and complicated, many consumers are confused. To give consumers a better understanding of Taiwanese tea, the Taiwan Tea Experiment Station (TTES) is going on in its attempts to formulate an evaluation system for tea quality. In general, there are four criteria used to evaluate tea quality: appearance, tincture, liquid color and aroma.

Because tea evaluation comes under the subjective judgment of experts, the quality levels are described using the terms: perfect, good, medium, poor and bad. These five quality defining terms allow for the ambiguity and imprecision inherent to human perception. Since fuzzy sets are

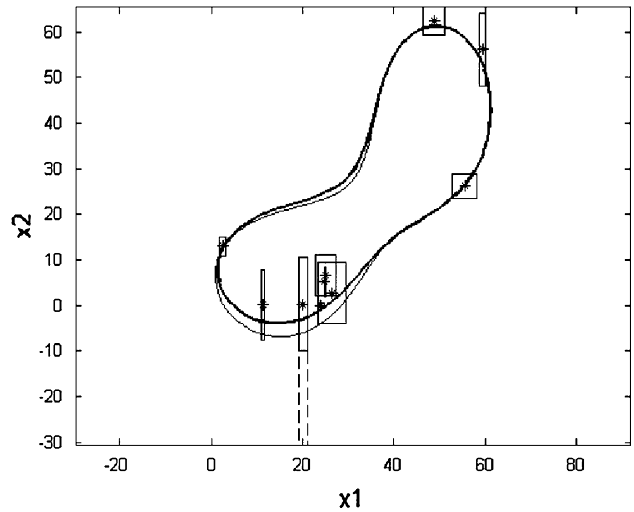

Fig. 7 *Thick curve* The description of data before changing the maximum width of uncertainty of second feature of the training sample (shown by *dashed rectangle*). *Thin curve* The description of data after changing the maximum width of uncertainty of the training sample

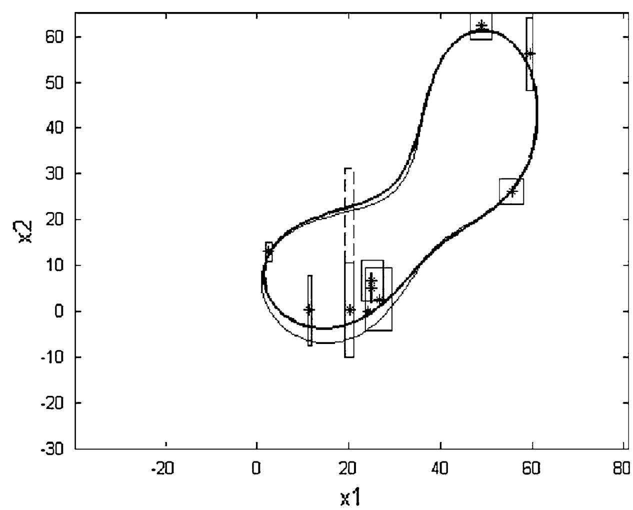

Fig. 8 *Thick curve* The description of data before changing the maximum width of uncertainty of second feature of the training sample (shown by *dashed rectangle*). *Thin curve* The description of data after changing the maximum width of uncertainty of the training sample

suited to describing ambiguity and imprecision in natural language, these terms can be defined using triangular fuzzy numbers as follows: $X_{\mathrm{perfect}} = (1, 0.25, 0)_{\mathrm{T}}$, $X_{\mathrm{good}} = (0.75, 0.25, 0.25)_{\mathrm{T}}$, $X_{\mathrm{medium}} = (0.5, 0.25, 0.25)_{\mathrm{T}}$, $X_{\mathrm{poor}} = (0.25, 0.25, 0.25)_{\mathrm{T}}$ and $X_{\mathrm{bad}} = (0, 0, 0.25)_{\mathrm{T}}$. These representations were shown in Fig. 11. Because tea evaluation will vary according to the evaluation of each individual expert, 10 experts were assigned to evaluate each kind of tea and assign the quality levels of perfect, good, medium, poor and bad for the four criteria of appearance, tincture,

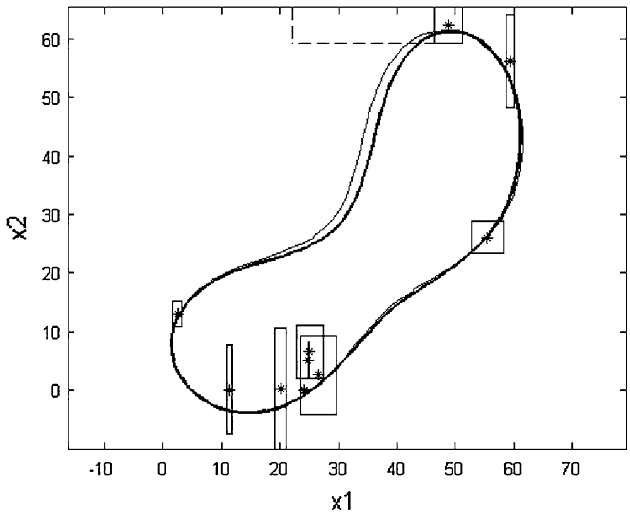

Fig. 9 *Thick curve* The description of data before changing the maximum width of uncertainty of first feature of the training sample (shown by *dashed rectangle*). *Thin curve* The description of data after changing the maximum width of uncertainty of the training sample

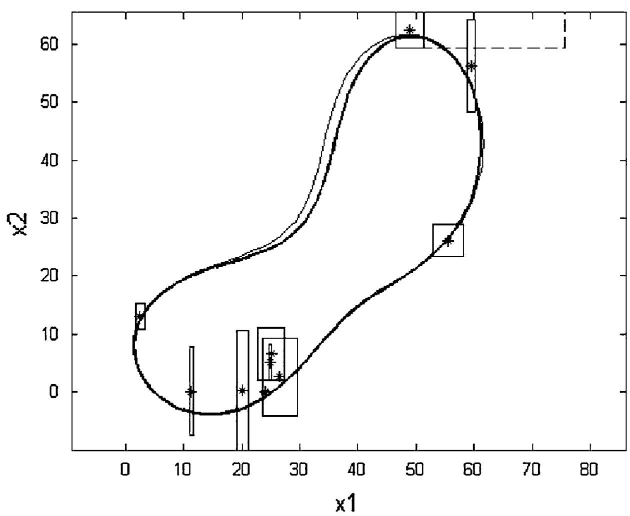

Fig. 10 *Thick curve* The description of data before changing the maximum width of uncertainty of first feature of the training sample (shown by *dashed rectangle*). *Thin curve* The description of data after changing the maximum width of uncertainty of the training sample

liquid color and aroma. For each criterion, a fuzzy arithmetic average was used to obtain a fuzzy number and then perfect, good, medium, poor or bad nearest was assigned to this fuzzy number. The final evaluation data were shown in Table 1. Let $\tilde{x}_{jk} = (m_{jk}, l_{jk}, r_{jk})_{\mathrm{T}}$ be assessed by kth criterion of the jth type of tea, $k = 1, 2, 3, 4, j = 1, \ldots, 69$. The overall performance for jth type of tea is determined as $\tilde{x}_j = (\bar{m}_j, \bar{l}_j, \bar{r}_j)_{\mathrm{T}}$ where

$$\bar{m}_j = \frac{1}{4}\sum_{k=1}^{4} m_{jk},\ \bar{l}_j = \frac{1}{4}\sum_{k=1}^{4} l_{jk},\ \bar{r}_j = \frac{1}{4}\sum_{k=1}^{4} r_{jk}. \quad (27)$$

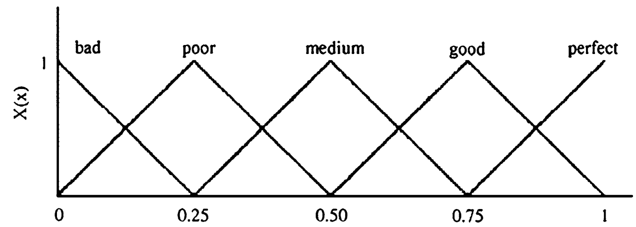

Fig. 11 Five triangular fuzzy numbers for a particular criterion

We clustered the dataset to two clusters using the FCN [23] and also the AFCN [24]. Data no. 1-54 and 55-69 were clustered to Grades 1 and 2, respectively. If we accept the result of these two clustering algorithms and use the first cluster of the dataset as the genuine class and the second cluster as the outlier, we can check the ability of our proposed method for one-class classification of the genuine class, namely Grade-one-tea.

In order to measure the probability of misclassification of our proposed method for one-class classification of the genuine class, we use the leave-one-out strategy (see e.g. [25, 26]), i.e., in turns, we remove only one element from the genuine class, we train the model with the remaining elements of the genuine class and we test this model with the sample removed from the genuine class and the elements of the outlier class. We repeat the process for every element of the genuine class and then obtain the probability of misclassification as follows:

$$\text{probability of misclassification} = \frac{\#\text{ of misclassifications}}{\#\text{ of test samples}}. \quad (28)$$

The obtained results for the mentioned strategy have been shown in Table 2. As it can be seen, the probability of misclassification is very low for different values of σ and the best result has been obtained for $\sigma = 0.02$.

We also used the 10-fold cross-validation to measure the probability of misclassification of our proposed method, i.e., the instances of the genuine class of the dataset are grouped in 10 sets (these sets forming a partition), and each one (together the outlier class) is used in turn as test set against all 9 others taken together as training set, i.e., the process is repeated 10 times and then obtain the probability of misclassification by using (28) (see [26]). The probability of misclassification of our proposed method on the dataset for 10-fold cross-validation strategy has been shown in Table 3. As it can be seen, however, the probability of misclassification of our proposed method for the 10-fold cross-validation strategy is more than its probability of misclassification for the leave-one-out strategy, but this probability is still small specially for $\sigma = 0.02$.

Table 1 69 types of tea tree in Taiwan

No.	Type	Appearance	Tincture	Liquid color	Aroma	$\bar{x}$
1	Bai Mao Hou	Perfect	Poor	Poor	Good	$(0.5625, 02500, 0.1875)_T$
2	Hei Mao Hou	Good	Poor	Poor	Good	$(0.5000, 0.2500, 0.2500)_T$
3	Qing Xin Hei Nou	Good	Poor	Poor	Good	$(0.5000, 0.2500, 0.2500)_T$
4	Qui Zi Keng Bai Mao	Good	Poor	Poor	Good	$(0.5000, 0.2500, 0.2500)_T$
5	Da Nan Wan Bai Mao	Perfect	Bad	Bad	Good	$(0.4375, 0.1250, 0.1875)_T$
6	Qing Xin Oolong	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
7	Dan Shui Qing Xin	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
8	Bu Zhi Chun	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
9	Tao Ren Chong	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
10	Wan Chong	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
11	Hong Xin Da Nou	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
12	Bai Xin Oolong	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
13	Shui Xian	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
14	Gui Hua Chong	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
15	Niu Pu Chong	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
16	Lin Kou Heng Zhe	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
17	Feng Zi Lin	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
18	Pu Xin Chong	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
19	Da Hu Wei	Good	Bad	Poor	Good	$(0.4375, 0.1875, 0.2500)_T$
20	Hong Xin Oolong	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
21	Fu Chou Chong	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
22	Tieh Kuan Yin	Medium	Bad	Bad	Perfect	$(0.3750, 0.1250, 0.1875)_T$
23	Heng Zhe Da Ye	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
24	Gan Zi Chong	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
25	San Cha Zhi Lan	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
26	Ying Zhi Hong Xin	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
27	Da Ye Oolong	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
28	Tian Gong Chong	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
29	Gan Zi Chong(Huang)	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
30	Wen Shen Da Ye	Good	Bad	Poor	Medium	$(0.3750, 0.1875, 0.2500)_T$
31	Hei Mian Zao Chong	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
32	Qing Xin Zao Chong	Medium	Bad	Poor	Good	$(0.3750, 0.1875, 0.2500)_T$
33	Lin Kou Da Ye	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
34	Zao Chong	Good	Bad	Bad	Good	$(0.3750, 0.1250, 0.2500)_T$
35	Han Kou Chong	Medium	Bad	Poor	Good	$(0.3750, 0.1875, 0.2500)_T$
36	Niu Shi Wu	Good	Bad	Poor	Medium	$(0.3750, 0.1875, 0.2500)_T$
37	Ping Shui Chong	Medium	Bad	Poor	Good	$(0.3750, 0.1875, 0.2500)_T$
38	Yan Chuan Chong	Good	Bad	Poor	Medium	$(0.3750, 0.1875, 0.2500)_T$
39	Da Ye Zhu Ye	Medium	Bad	Bad	Good	$(0.3125, 0.1250, 0.2500)_T$
40	Tao Ren Wu	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
41	Hu Nan Chong	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
42	Huang Zhi Chong	Medium	Bad	Bad	Good	$(0.3125, 0.1250, 0.2500)_T$
43	Ji Long Jin Gui	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
44	Wu Jin Chong	Medium	Bad	Bad	Good	$(0.3125, 0.1250, 0.2500)_T$
45	Jin Gui Chong	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
46	Da Ji Ling Chong	Medium	Bad	Poor	Medium	$(0.3125, 0.1850, 0.2500)_T$
47	Huang Gan	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
48	Zhi Lan Chong	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$

Table 1 continued

No.	Type	Appearance	Tincture	Liquid color	Aroma	$\bar{x}$
49	Shi Tea	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
50	Bai Xin Wu Yi	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
51	Mao Er Chong	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
52	Ji Long Bai Chong	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
53	Zhu Ye Chong	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
54	Yu Zhi Chong	Good	Bad	Bad	Medium	$(0.3125, 0.1250, 0.2500)_T$
55	Xiao Ye Zhu Ye	Medium	Bad	Bad	Medium	$(0.2500, 0.1250, 0.2500)_T$
56	Shen Man Chong	Medium	Bad	Bad	Medium	$(0.2500, 0.1250, 0.2500)_T$
57	Bai Chong	Medium	Bad	Bad	Medium	$(0.2500, 0.1250, 0.2500)_T$
58	Bai Ye Chong	Medium	Bad	Bad	Medium	$(0.2500, 0.1250, 0.2500)_T$
59	Yellow Tea	Medium	Bad	Bad	Medium	$(0.2500, 0.1250, 0.2500)_T$
60	Manipuri	Good	Bad	Bad	Bad	$(0.1875, 0.0625, 0.2500)_T$
61	Shan	Good	Bad	Bad	Bad	$(0.1875, 0.0625, 0.2500)_T$
62	Gao Lu Chong	Medium	Bad	Bad	Bad	$(0.1250, 0.0625, 0.2500)_T$
63	Indigenou	Medium	Bad	Bad	Bad	$(0.1250, 0.0625, 0.2500)_T$
64	Nan Tou Shen Tea	Medium	Bad	Bad	Bad	$(0.1250, 0.0625, 0.2500)_T$
65	Japuri	Medium	Bad	Bad	Bad	$(0.1250, 0.0625, 0.2500)_T$
66	A Sa Mu	Medium	Bad	Bad	Bad	$(0.1250, 0.0625, 0.2500)_T$
67	Mian Dian Chong	Medium	Bad	Bad	Bad	$(0.1250, 0.0625, 0.2500)_T$
68	Shan Tea	Medium	Bad	Bad	Bad	$(0.1250, 0.0625, 0.2500)_T$
69	Kyang	Medium	Bad	Bad	Bad	$(0.1250, 0.0625, 0.2500)_T$

Table 2 The probability of misclassification of our proposed method for leave-one-out strategy

σ	0.02	0.04	0.06	0.08	0.1	0.2	0.3	0.4	0.5
Probability of misclassifications	0.0278	0.0313	0.0313	0.0324	0.0359	0.0336	0.0336	0.0324	0.0313

Table 3 The probability of misclassification of our proposed method for 10-fold cross validation strategy

Σ	0.02	0.04	0.06	0.08	0.1	0.2	0.3	0.4	0.5
Probability of misclassification	0.0918	0.2102	0.1416	0.1827	0.1504	0.1645	0.1329	0.0920	0.2238

5 Conclusion

In this paper, a fuzzy quadratic program with fuzzy quadratic objective function and quadratic fuzzy constraints was solved and this solution was used to obtain the description of fuzzy data based on the FSVDD method. The proposed method, namely FSVDD*, is suitable for one-class classification of real data which are usually uncertain.

We used a defuzzification method to solve our fuzzy quadratic program. The defuzzification of fuzzy parameters of the fuzzy quadratic program leads to obtain a crisp quadratic program which gets a crisp description of fuzzy data. We used the Hathaway, the Hausdorff, and the Yang distance for defuzzification and studied the advantage and drawback of each distance metric on the fuzzy data description. Finally, we applied our proposed method to real data. The experimental results showed the ability of the proposed method in Taiwanese tea evaluation.

Appendix

The Lagrangian dual form of the program (1) is as follows:

$$\max_{\delta,\gamma} L(R, e, \xi, \delta, \gamma) \quad \text{subject to } \delta_i, \gamma_i \geq 0, \quad i = 1, \ldots, n, \tag{29}$$

where $\delta = (\delta_1, \ldots, \delta_n)^{\mathrm{T}}$, $\gamma = (\gamma_1, \ldots, \gamma_2)^{\mathrm{T}}$ and

$$L(R, e, \xi, \delta, \gamma) = \inf\left\{R^2 + C\sum_{i=1}^{n} w_i\xi_i - \sum_{i=1}^{n}\delta_i(R^2 + \xi_i - g(x_i)^{\mathrm{T}}g(x_i) + 2e^{\mathrm{T}}g(x_i) - e^{\mathrm{T}}e) - \sum_{i=1}^{n}\gamma_i\xi_i\right\}. \tag{30}$$

For the optimal solution, the following conditions are satisfied

$$\frac{\partial L}{\partial R} = 0 \rightarrow \sum_{i=1}^{n}\delta_i = 1, \tag{31}$$

$$\frac{\partial L}{\partial e} = 0 \rightarrow e = \sum_{i=1}^{n}\delta_i g(x_i), \tag{32}$$

$$\frac{\partial L}{\partial \xi} = 0 \rightarrow \delta_i = Cw_i - \gamma_i, \quad i = 1, \ldots, n, \tag{33}$$

$$\delta_i(\|g(x_i) - e\|^2 - R^2 - \xi_i) = 0, \quad i = 1, \ldots, n, \tag{34}$$

$$\gamma_i\xi_i = 0, \quad i = 1, \ldots, n, \tag{35}$$

Using the above conditions, $L(R, e, \xi, \delta, \gamma)$ is transformed to

$$\sum_{i=1}^{n}\delta_i K(x_i, x_i) - \sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i\delta_j K(x_i, x_j), \tag{36}$$

where $K(x_i, x_j) = g(x_i)^{\mathrm{T}}g(x_j)$. Since $\delta_i \geq 0$ and from (33) we have $0 \leq \delta_i \leq Cw_i$. So, the Lagrangian dual form of (1) can be restated as follows:

$$\max_{\delta} \sum_{i=1}^{n}\delta_i K(x_i, x_i) - \sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i\delta_j K(x_i, x_j) \quad \text{subject to} \begin{cases} \sum_{i=1}^{n}\delta_i = 1, \\ 0 \leq \delta_i \leq Cw_i, \quad i = 1, \ldots, n, \end{cases} \tag{37}$$

which is a conventional quadratic program and can be solved easily. From (32), $\|g(x_i) - e\|^2 = K(x_i, x_i) - 2\sum_{j=1}^{n}\delta_j K(x_i, x_j) + \sum_{j=1}^{n}\sum_{k=1}^{n}\delta_j\delta_k K(x_j, x_k)$ and from (34) if $\delta_i > 0$, $K(x_i, x_i) - 2\sum_{j=1}^{n}\delta_j K(x_i, x_j) + \sum_{j=1}^{n}\sum_{k=1}^{n}\delta_j\delta_k K(x_j, x_k) = R^2 + \xi_i$. From (33) if $\delta_i < Cw_i$, $\gamma_i > 0$. So, from (35) we have $\xi_i = 0$. So, if $0 < \delta_i < Cw_i$,

$$R^2 = K(x_i, x_i) - 2\sum_{j=1}^{n}\delta_j K(x_i, x_j) + \sum_{j=1}^{n}\sum_{k=1}^{n}\delta_j\delta_k K(x_j, x_k). \tag{38}$$

Finally, the unknown datum x is inside the hypersphere if $\|g(x) - e\|^2 \leq R^2$ or equivalently if

$$K(x, x) - 2\sum_{i=1}^{n}\delta_i K(x, x_i) + \sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i\delta_j K(x_i, x_j) \leq R^2. \tag{39}$$

References

1. Bigand A, Colot O (2010) Fuzzy filter based on interval-valued fuzzy sets for image filtering. Fuzzy Sets Syst 161:96–117
2. Hong D, Hwang C (2006) Fuzzy nonlinear regression model based on LS-SVM in feature space. In: Fuzzy Systems and Knowledge Discovery, pp 208–216
3. Yang R, Wang Z, Heng P-A, Leung K-S (2007) Classification of heterogeneous fuzzy data by Choquet integral with fuzzy-valued integrand. IEEE Trans Fuzzy Syst 15:931–942
4. Hao P-Y (2008) Fuzzy one-class support vector machines. Fuzzy Sets Syst 159:2317–2336
5. Gambardella A, Giacinto G, Migliaccio M, Montali A (2009) One-class classification for oil spill detection. Pattern Anal Appl 13:349–366
6. Cordella L, Sansone C (2007) A multi-stage classification system for detecting intrusions in computer networks. Pattern Anal Appl 10:83–100
7. Khuwaja G, Abu-Rezq A (2004) Bimodal breast cancer classification system. Pattern Anal Appl 7:235–242
8. Rajanna U, Erol A, Bebis G (2010) A comparative study on feature extraction for fingerprint classification and performance improvements using rank-level fusion. Pattern Anal Appl 13:263–272
9. Torkkola K (2004) Discriminative features for text document classification. Pattern Anal Appl 6:301–308
10. Song F, Liu S, Yang J (2005) A comparative study on text representation schemes in text categorization. Pattern Anal Appl 8:199–209
11. Shanthi N, Duraiswamy K (2009) A novel SVM-based handwritten Tamil character recognition system. Pattern Anal Appl 13:173–180
12. Bajla I, Holländer I, Czedik-Heiss D, Granec R (2009) Classification of image objects in Epo doping control using fuzzy decision tree. Pattern Anal Appl 12:285–300
13. Khuwaja G (2005) Merging face and finger images for human identification. Pattern Anal Appl 8:188–198
14. Tolba AS, Abu-Rezq AN (2000) Combined classifiers for invariant face recognition. Pattern Anal Appl 3:289–302
15. Sansone C, Vento M (2000) Signature verification: increasing performance by a multi-stage system. Pattern Anal Appl 3:169–181
16. Tax DMJ, Duin RPW (1999) Support vector domain description. Pattern Recogn Lett 20:1191–1199
17. Tax DMJ, Duin RPW (2004) Support vector data description. Mach Learn 54:45–66
18. Zhang Y, Chi Z-X, Li K-Q (2009) Fuzzy multi-class classifier based on support vector data description and improved PCM. Expert Syst Appl 36:8714–8718

19. Bazara MS, Sherali HD, Shetty CM (2006) Nonlinear programming, 3rd edn. Wiley, New York
20. Liu S-T (2009) A revisit to quadratic programming with fuzzy parameters. Chaos Solitons Fractals 41:1401–1407
21. Tian J, Ha M-H, Li J-H, Tian D-Z (1996) The fuzzy-number based key theorem of statistical learning theory. In: Machine learning and cybernetics, Dalian, China, pp 3475–3479
22. Hathaway RJ, Bezdek JC, Pedrycz W (1996) A nonparametric model for fusing heterogeneous fuzzy data. IEEE Trans Fuzzy Syst 4:270–281
23. Yang M-S, Ko C-H (1996) On a class of fuzzy c-numbers clustering procedures for fuzzy data. Fuzzy Sets Syst 84: 49–60
24. Hung W-L, Yang M-S (2005) Fuzzy clustering on LR-type fuzzy numbers with an application in Taiwanese tea evaluation. Fuzzy Sets Syst 150:561–577
25. Hand DJ, Mannila H, Smyth P (2001) Principles of data mining. MIT Press, Cambridge
26. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on artificial intelligence, pp 1137–1143