Fuzzy LSSVC-WKNN Combination Algorithm in Fault Diagnosis

Sheng Sun[1], Chuiwei Lu[1], Guojun Zhang[1]

[1]Computer School, Hubei Polytechnic University, Huangshi, Hubei, China

Abstract

In this paper, we propose a fuzzy least squares support vector clustering-weighted k-nearest neighbour (LSSVC-WKNN) combination algorithm. We use the ATS method to optimize feature vectors and adopt this algorithm for fault diagnosis. Firstly, the fault diagnosis model is described, and then the feature vector is reduced by ATS method. The fuzzy LSSVC algorithm is adopted to train the optimized feature vector of training samples, so as to obtain a trained clustering model. Then, the fuzzy LSSVC-WKNN combination algorithm is used for fault diagnosis of test samples. Taking the simulated circuit fault diagnosis test as an example, the experimental results prove that the method in this paper has the advantage of higher diagnostic accuracy than other methods, and it is a universal and feasible online fault diagnosis method.

Keywords: least squares, support vector, clustering, k-nearest neighbour, fault diagnosis

Received: 21 November 2018

To cite this article:
Sun Sheng, Lu Chuiwei, Zhang Guojun, "Fuzzy LSSVC-WKNN combination algorithm in fault diagnosis" in *Electrotehnica, Electronica, Automatica (EEA)*, 2020, vol. 68, no. 2, pp. 93-101, ISSN 1582-5175.

1. Introduction

KNN algorithm is the most usable classification algorithm, it is simple, straight and effective. Unlike SVM or decision tree algorithm, KNN algorithm is a kind of typical passive learning algorithm, any classification model need not to set up basically in the training stage, records only the distance between samples, in the classification process, KNN calculate the distance between the samples to be classified and known data, choose a certain number of known neighbour samples, The class with the largest number of samples is taken as the class of samples to be classified. KNN algorithm does not need complicated statistical modelling technology, but only needs to calculate the number of known samples of different categories in a certain neighbourhood and can be directly applied to multiple classification problems.

KNN algorithm has significant deficiency: category balance problem. Category balancing problem is also known as category sensitive problem, which means when the number of training samples belonging to different categories varies greatly, it will greatly affect the classification performance of the algorithm. When the data belonging to a certain category of training samples are significantly more than other categories, due to the accidental factors, the probability of falling into the neighbourhood of the samples to be classified increases greatly, which may affect the final category of the data points to be classified. In order to overcome the problem of category imbalance, we select the training samples with the same number of other categories from the category with the most obvious training data. KNN cannot identify the effect of attributes in dataset, a major approach to tackle this problem is to give each of attributes a weight value according to the relationship between these attributes. The bigger the attribute weight is, it has more importance extent in figuring out the distance of samples. Thus the reliability and classification accuracy of training can be improved at the cost of less time.

Support vector clustering (SVC) originated from the concept of support vector machines has recently drawn the attention of clustering research communities because of its effectiveness in data clustering [1]. Kernel Methods [2]-[4] are algorithms that, by replacing the inner product with an appropriate positive definite function, implicitly perform a nonlinear mapping of the input data to a high dimensional feature space. LS-SVC consists of using equality constraints at the primal cost function to be minimized, instead of inequality ones. Moreover, the 2-norm is used for the slack variables of the primal problem, instead of the 1-norm. As a result of these modifications, the problem generated by LS-SVC can be solved with a system of linear equations, which is less complex than the quadratic programming used in SVC. LS-SVC has many advantages: the global optimal solution can be obtained by solving quadratic programming problem [5]-[7]. And this method can deal with huge amounts of data, high dimensional data and noise [12]-[17].

The drawback of the LS-SVC method is the loss of sparseness [18]-[21]. In order to overcome the drawback within the LS-SVC framework, we apply fuzzy least squares in order to acquire robustness, a support degree is defined for each training point, the support degree corresponding to training data represents the data belonging to the degree of support vector, and the support degree is determined by the absolute value of Lagrangian multipliers.

Fault Diagnosis Technology can be regarded as the science of equipment and system fault detection, separation and identification through the comprehensive utilization of system historical state information and current state through some analysis methods [22]-[24]. Fault diagnosis technology has developed rapidly in the past decades. The existing fault diagnosis methods can be divided into: model-based method, knowledge reasoning and intelligent technology, signal extraction and processing method, clustering method.

For model-based fault diagnosis approaches, a system model, which explicitly describes the relationship among the system variables, is available to the designer. Based on the model, fault diagnosis schemes/algorithms can be designed and then implemented online for monitoring and diagnosing the real-time system/process. In model-based methods, the models of the industrial processes or the practical systems are required to be available, which can be obtained by using either physical principles or systems identification techniques. Based on the model, fault diagnosis algorithms are developed to monitor the consistency between the measured outputs of the practical systems and the model-predicted outputs.

For signal-based fault diagnosis methods, the signal pattern/symptom of a system under a healthy status is a priori, and the fault diagnosis is carried out by checking the consistency between the known healthy signal pattern and the signal symptom of the real-time process extracted either by using time-domain, frequency-domain, or time-frequency signal processing techniques.

Signal-based methods utilize measured signals rather than explicit input-output models for fault diagnosis. The faults in the process are reflected in measured signals, whose features are extracted, and a diagnostic decision is then made based on the symptom analysis and prior knowledge on the symptoms of the healthy systems. Signal-based fault diagnosis methods have a wide application in real-time monitoring and diagnosis for induction motors, power converters, and mechanical components in a system.

For complicated industrial processes, a large amount of historical data, rather than a model or a signal pattern, is available. The underlying knowledge, which implicitly represents the dependence of the system variables, can be extracted by using various artificial intelligent techniques and the available historic data. Fault diagnosis is carried out by checking the consistency of the obtained underlying knowledge and the real-time system feature extracted from the online monitored data. Different from model-based methods and signal-based approaches that require either an a priori known model or signal patterns, knowledge-based fault diagnosis methods start from where only a large volume of historic data is available. Applying a variety of artificial intelligent techniques to the available historic data of the industrial processes, the underlying knowledge, which implicitly represents the dependence of the system variables, can be extracted. The consistency between the observed behaviour of the operating system and the knowledge base is then checked, leading to a fault diagnosis decision with the aid of a classifier.

Clustering is an important data-driven approach in fault diagnosis. It does not need accurate mathematical models and can use a large amount of system historical and online data to extract failure information and to realize fault diagnosis. However, existing methods and technologies cannot handle new faults well.

The above work is of great significance, but the rationality of the model is difficult to be guaranteed, the expert system needs prior knowledge, new faults cannot be handled well and the diagnosis efficiency is not high. Therefore, a fuzzy LSSVC-WKNN combination diagnosis algorithm is designed in this paper.

The rest of this paper is organized as follows. The k-nearest neighbour clustering algorithm will be described in section 2. Section 3 presents a weighted k-nearest neighbour clustering algorithm. Section 4 presents a fuzzy least squares support vector clustering algorithm. Section 5 introduces our fuzzy LSSVC-WKNN combination algorithm and use this method to solve the problem of fault detection. Section 6 discusses results and section 7 concludes.

2. K-Nearest Neighbour Clustering Algorithm

KNN is an extension of 1nn, its thought is simple: For the following clustering problem with *N* samples, suppose these samples have been partitioned into *c* clusters $\omega_i(i=1,2,\cdots,c)$, and cluster ω_i includes N_i samples $x_j^{(i)}(j=1,2,\cdots,N_i)$. When given a new unidentified sample *x*, KNN clustering algorithm searches the *k* training samples that are nearest to the new sample *x* as the new sample neighbours, which cluster includes most of these *k* samples, then predict the cluster label of *x* is same with this cluster. If the number of samples that the new sample neighbours belong to cluster $\omega_i(i=1,2,\cdots,c)$ is $k_1,k_2,\cdots,k_c$ respectively, we can define decision function of ω_i is:

$$d_i(x)=k_i, \quad i=1,2,\cdots,c \tag{1}$$

Decision rule is:

If $d_m(x)=\max_{i=1,2,\cdots,c} d_i(x)$, then predict $x\in\omega_m$.

This method is K-nearest neighbour clustering algorithm.

When $N\to\infty$, the error fraction of KNN satisfies the following inequality:

$$\hat{P}\le P_e\le\hat{P}[2-\hat{P}c/(c-1)] \tag{2}$$

Where, $\hat{P}$ is the error score of Bayesian classification rate, and c is the number of categories.

When the sample number N is limited, the following distance definition can guarantee the minimum mean square error with the KNN error fraction rate of infinite sample:

$$D(x,x^L) = \left|\nabla P(\omega_1|x)^T (x - x^L)\right| \quad (3)$$

KNN is applicable on the premise that the number of samples is large, and the conditional probability of local neighbour region is the same. For the selection of k value, on the one hand, a larger k value can reduce the error score rate, and on the other hand, k neighbours are required to be close to the test sample. The drawback of KNN is that both the amount of calculation and the amount of storage are large.

3. The Weighted K-Nearest Neighbour Clustering Algorithm

The specific implementation process of the weighted KNN algorithm proposed in this paper is as follows:

(1) Evaluate the weight of the property.

Through the importance of attributes, the weight of each attribute can be calculated.

Suppose original attribute set is $F = \{F_1, F_2, \cdots, F_n\}$, the total amount of attributes are n, $dis(F_i, F_j)$ means the distant between F_i and F_j, FS is attribute set which has eliminated redundant attributes. We propose an ATS algorithm to get *m* attributes which have greater effect on clustering. Its basis principle is: Test the effect degree of attributes on clustering results, hold the important attributes which have greater effect on clustering, and discards the attributes which are less effective. The ATS method of attribute selection is as follows:

Algorithm 1: ATS (attribute selection)
Input:
The original attribute set *F*;
Output:
Selected attribute set *FS*, which has eliminated redundant attributes;
Redundant attribute set *FR*;
Method:
Step 1: initialization: $FS \leftarrow F$;
Step 2: in the sample set which includes all n attributes, cluster each sample, and compute cluster error ε_0;
Step 3: for $i = 1$ to n do
discard F_i in *FS*, that is, $FS = FS - \{F_i\}$, then compute cluster error ε_i;
Step 4: sort ε_i in descending order:
$\varepsilon_1' > \varepsilon_2' > \cdots > \varepsilon_{n-1}' > \varepsilon_n'$;
Step 5: if $\varepsilon_n' < \varepsilon_0$
then discard F_n' which correspond to ε_n' in *FR*, that is, $FS = FS - \{F_n'\}$;
$n := n-1$; $\varepsilon_0 := \varepsilon_n'$; goto Step 3
else goto (g);
Step 6: until $n < m+1$ (m is some fixed value);
Step 7: output *FS* and *FR*, $FS = \{F_1^*, F_2^*, \cdots, F_m^*\}$,
$FR = F - FS = \{F_{m+1}^*, F_{m+2}^*, \cdots, F_n^*\}$,
Stop the algorithm.

Through ATS function, we can get m attributes which have greater effect on clustering. These m attributes still have different effect degree on results, so we set each attribute in *FS* a weight value $w_j (j = 1, 2, \cdots, m)$. The bigger the weight value is, the attribute affect more in figuring out the distance.

We can calculate the weight of each attribute according to the following formula:

$$w_j = \frac{dis(F_j^*, FR)}{\max\limits_{i=1,2,\cdots,m} dis(F_i^*, FR) - dis(F_j^*, FR)} \quad (4)$$

$$dis(F_j^*, FR) = \frac{\sum\limits_{q=m+1}^{n} dis(F_j^*, F_q^*)}{n-m} \quad (5)$$

(2) Find the weighted distance between the test sample and all samples of the original training set.

Aiming at the disadvantage that Euclidean distance does not show the importance of attribute, the weighted Euclidean distance is proposed. The weighted Euclidean distance between sample $x_1 = [x_{11}, \cdots, x_{1m}]$ and sample $x_2 = [x_{21}, \cdots, x_{2m}]$ is:

$$d(x_1, x_2) = \left(\sum_{j=1}^{m} w_j (x_{1j} - x_{2j})^2\right)^{1/2} \quad (6)$$

where, m represents the number of attributes after reduction, w_j is the weight of the jTH attribute, and $w_j > 0, \sum_{j=1}^{m} w_j = 1$.

Through (7), the Euclidean distance between the test sample x and all samples of the training set is calculated and weighted to obtain the final distance.

$$d_i(x) = \left(\sum_{j=1}^{n} w_j (x_j - t_{ij})^2\right)^{1/2} \quad (7)$$

where, $i = 1, 2, \cdots, N$, $j = 1, 2, \cdots, m$,

N represents the number of training samples, x_j represents the jTH conditional attribute value of the test sample, t_{ij} represents the jTH conditional attribute value of training sample i.

(3) According to equation (7), k training samples with the minimum distance from the test sample are obtained and the value of k is set. After finding the weighted distance d between the test sample and all

samples in the training set by equation (7), k nearest neighbours of the test sample in the training sample are selected according to the principle of minimum distance.

(4) After obtaining k neighbours through the previous step, the number of occurrences of all kinds of samples is counted, and the class with the most number is calculated. The class of test sample is classified as that class. If the samples of different classes have the same number of occurrences in k neighbours and all are the most, then go back to the previous step and continue to find the training samples with the minimum distance. Only when the class with the largest training samples can be counted can the training samples be withdrawn from the circulation, and the test samples are judged as this class.

4. The Fuzzy Least Squares Support Vector Clustering Algorithm

In the paper [10], we have given the mathematical formulation of the LSSVC algorithm. In order to achieve robustness, each error variable ξ_i is given a weight, assume the weight corresponds to ξ_i is v_i. The objective function of the fuzzy LSSVC is described as the following optimization problem

$$\min_{R,a,\xi_i} R^2 + \frac{1}{2}C\sum_i v_i\xi_i^2 \quad (8)$$

Subject to

$$\|\varphi(x_i) - a\|^2 = R^2 + \xi_i \quad (9)$$

where R is the radius and the centre of the enclosing sphere is the point a; ξ_i is a error variable; $C>0$ is the penalty parameter, which is a constant that determines the trade-off between the volume of the hypersphere and the number of the outliers. In LSSVC, because all the Lagrangian multipliers β_i are not zero, so all the data vectors are support vectors [8]-[11]. We define a support degree value for each training data. The support degree corresponding to training data (x_i, y_i) has the value $0 < v_i < 1$, represents the data belonging to the degree of support vector. The greater the value of v_i, the higher the corresponding training point belonging to the degree of support vector. To solve this problem, we introduce the Lagrangian:

$$L(R,a,\xi_i,\beta_i) = R^2 + \frac{1}{2}C\sum_i v_i\xi_i^2 - \sum_i (R^2 + \xi_i - \|\phi(x_i) - a\|^2)\beta_i \quad (10)$$

With Lagrange multipliers (called support values). The conditions for optimality are given by

$$\frac{\partial L(R,a,\xi_i,\beta_i)}{\partial R} = 2R - 2R\sum_i \beta_i = 0 \Rightarrow \sum_i \beta_i = 1 \quad (11)$$

$$\frac{\partial L(R,a,\xi_i,\beta_i)}{\partial a} = -\sum_i 2\beta_i(\phi(x_i) - a) = 0 \Rightarrow a = \sum_i \beta_i\phi(x_i) \quad (12)$$

$$\frac{\partial L(R,a,\xi_i,\beta_i)}{\partial \xi_i} = Cv_i\xi_i - \beta_i = 0 \Rightarrow \beta_i = Cv_i\xi_i \quad (13)$$

Now substitute (11) to (13) into equation (10), we get the dual problem of (8) as follows:

$$\min_{\beta_i} W = \frac{1}{2C}\sum_i \frac{\beta_i^2}{v_i} + \sum_i\sum_j \beta_i\beta_j\phi(x_i)\cdot\phi(x_j) - \sum_i \phi(x_i)^2\beta_i \quad (14)$$

subject to the constraints $\beta_i > 0, \sum_i \beta_i = 1, i = 1,\cdots,N$.

and the Karush-Kuhn-Tucker conditions for $i = 1,\cdots,N$

$$(R^2 + \xi_i - \|\phi(x_i) - a\|^2)\beta_i = 0 \quad (15)$$

which $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $k(\mathbf{x},\mathbf{y})$ is kernel function. We rewrite equation (14) to the standard quadratic form

$$\min f(\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\beta}^T\mathbf{Q}\boldsymbol{\beta} - \mathbf{P}^T\boldsymbol{\beta} \quad (16)$$

Subject to

$$\mathbf{I}^T\boldsymbol{\beta} = 1 \quad (17)$$

which $\mathbf{Q} = C^{-1}\mathbf{V} + 2\mathbf{K}$, P is N dimension column vector, $\mathbf{P}^T = (k_{11}, k_{22}, \cdots, k_{NN})$, E is unit diagonal matrix, $\mathbf{V}$ is diagonal matrix, which the value of element on the principal diagonal is v_i, that is $\mathbf{V} = \mathrm{diag}\{v_1, v_2, \cdots, v_N\}$. $\mathbf{K}$ is $N \times N$ symmetric matrix, $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $\boldsymbol{\beta}$ is N dimension column vector, $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \cdots, \beta_N)$, I is N dimension unit column vector.

We can prove that (16) is a convex quadratic programming problem. Because for any $\boldsymbol{\beta} \in R^N$,

$$\boldsymbol{\beta}^T\mathbf{Q}\boldsymbol{\beta} = \boldsymbol{\beta}^T(\mathbf{v}^{-1}C^{-1}\mathbf{E} + 2\mathbf{K})\boldsymbol{\beta} = \frac{1}{C}\sum_i \frac{\beta_i^2}{v_i} + 2\sum_i\sum_j \beta_i\beta_j k_{ij}$$

$$= \frac{1}{C}\sum_i \frac{\beta_i^2}{v_i} + 2a^2 \geq 0 \quad (18)$$

So, $\mathbf{Q}$ is a positive semi-definite matrix, thus we can conclude that (16) is a convex quadratic programming problem.

After elimination of ξ_i, we can obtain the solution

$$\left[\begin{array}{c|c} 0 & \mathbf{1}^T \\ \hline \mathbf{1} & U + \Omega_{ii} - \beta^T\Omega_{ij} \end{array}\right]\left[\begin{array}{c} 0 \\ \hline \beta \end{array}\right] = \left[\begin{array}{c} 1 \\ \hline f \end{array}\right] \quad (19)$$

with $\mathbf{1} = [1;\cdots;1]$, $\beta = [\beta_1;\cdots;\beta_N]^T$, $f = [f_1;\cdots;f_N]^T$, and Ω is kernel matrix,

$$\Omega_{ij} = \varphi(x_i)^T \cdot \varphi(x_j) = \mathrm{K}(x_i, x_j) \quad i, j = 1, 2, \cdots, N \quad (20)$$

where the diagonal matrix U is given by

$$\mathrm{U} = \mathrm{diag}\{\frac{1}{\mathrm{C}v_1}, \frac{1}{\mathrm{C}v_2}, \cdots, \frac{1}{\mathrm{C}v_N}\} \tag{21}$$

The radius of the enclosing sphere can be computed by (22)

$$R^2 = \|\varphi(x_i) - \boldsymbol{a}\|^2 = \left| K(x_i, x_i) - 2\sum_i \beta_i K(x_i, x_i) + \sum_{i,j} \beta_i \beta_j K(x_i, x_j) \right|$$
$$= \sum_i \beta_i K(x_i, x_i) - \sum_{i,j} \beta_i \beta_j K(x_i, x_j) \tag{22}$$

where data points x_i lying on the surface of the sphere in the feature space.

The data at the moment i is given different reliability through the weighting factor, and its weighting rule should satisfy such a rule, that is, the weighting of each observation data should be inversely proportional to the observation error. This method can make the decision surface accurately separate one kind of sample from the whole sample. If this rule is combined with fuzzy weighting to make the membership function a linear function of error, it can be selected:

$$v_i = \begin{cases} 1 & if \ |\xi_i| < c_1 \\ -a|\xi_i| + d & if \ c_1 \le |\xi_i| \le c_2 \\ 10^{-4} & otherwize \end{cases} \tag{23}$$

The values of a and d can be obtained by using boundary conditions:

$$a = \frac{1-\sigma}{c_2 - c_1}, \ d = \frac{c_2 - c_1\sigma}{c_2 - c_1} \tag{24}$$

The value of σ is usually 10^{-4}.

The fuzzy weight function v_i is determined by the error variable ξ_i, where ξ_i is a robust estimate of the standard deviation of the LSSVC error variables, it is the observation error.

General process of our fuzzy LSSVC can be summarized as follows:

(1) Given training data $\{x_i, y_i\}_{i=1}^N$, find an optimal (C,σ) combination (The optimal selection of parameters (C,σ) can be determined by the two layers of grid search strategy or 10-fold cross-validation). The LS-SVC model is obtained by training process and we compute $\xi_i = \beta_i / C$.

(2) Determine the weights v_i based upon ξ_i.

(3) Solve the fuzzy LSSVC (19), obtain the values of β, and give the above model

$$f(x_i) = d^2(x_i) = \|\varphi(x_i) - \boldsymbol{a}\|^2 = \sum_i \beta_i K(x_i, x_i) - \sum_{i,j} \beta_i \beta_j K(x_i, x_j)$$

5. Fault Diagnosis Based On The Fuzzy Lssvc-Wknn Algorithm

5.1 Fault Diagnosis Model

The fault diagnosis model designed in this paper can be described as extracting feature signals, implementing feature optimization of fault signs by using ATS method, and realizing fault diagnosis by our fuzzy LSSVC-WKNN combination algorithm.

The overall structure of our fault diagnosis model is shown in figure 1.

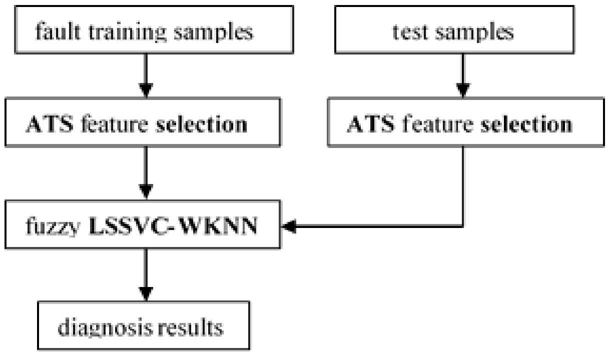

Figure 1. The chart of our fault diagnosis model

The specific process can be divided into the following four stages:

(1) extract fault symptom samples.

Fault symptom samples are presented in different devices and systems in different forms. In WSN, the combined vector composed of data collected by sensors and data collected by neighbouring nodes is used as fault diagnosis data.

(2) Adopt the ATS method to optimize the characteristics of the existing fault symptom samples.

The ATS method is used to reduce the dimension of fault symptom data, so the complexity of fault diagnosis data is reduced. The algorithm description is shown in algorithm 1.

(3) Fuzzy LSSVC clustering algorithm is adopted to achieve fault sample clustering.

The fuzzy LSSVC algorithm is used to cluster the optimized feature vectors of fault signs, and the model of fault diagnosis is obtained. The description of the fuzzy LSSVC algorithm is shown in section 4.

(4) The fuzzy LSSVC-WKNN algorithm is used to realize fault diagnosis.

5.2 The Fuzzy LSSVC-WKNN Method

In the practical application, we analysed the distribution of the error-clustering samples in the LSSVC algorithm and found that the error sample points of the LSSVC clustering were all near the hypersphere. This algorithm is easy to identify the test samples incorrectly near the hypersphere. To solve this problem, the information provided by samples near the interface needs to be utilized as much as possible to improve clustering performance. The support vectors with large weights are mostly located near the hypersphere, the sample points near the hypersphere are more important support vectors.

In this paper, a fuzzy LSSVC-WKNN combination algorithm is proposed. Its main idea is:

When the test sample is closer to the LSSVC hypersphere, that is, the distance between it and the hypersphere is less than a certain value, then the fuzzy LSSVC is used for clustering. When the distance between the test sample and the LSSVC hypersphere is greater than a certain value, that is, it is far from the hypersphere, then, a part of support vector with large weight is used as a neighbour sample of the test sample to carry out WKNN clustering.

Our fuzzy LSSVC-WKNN combination algorithm makes full use of the support vector with large weight to represent the meaning of the training data set. At the same time, the WKNN algorithm can reduce the dependence on the neighbour parameters and improve the clustering accuracy of the cluster.

The specific fuzzy LSSVC-WKNN algorithm is as follows:

Firstly, the fuzzy LSSVC algorithm is used to calculate the corresponding support vector and its fuzzy weight v_i. Set T as the test set, T_{sv} as the support vector set with its fuzzy weight $v_i > 0.8$, T_{nm} as the normal sample, T_{fl} as the fault sample set, and k as the number of KNN.

Step 1: If $T \neq \Phi$, then select $x_i \in T$; If $T = \Phi$, then stop the algorithm.

Step 2: Calculate the formula

$$f(x_i) = d^2(x_i) = \|\varphi(x_i) - \boldsymbol{a}\|^2 = \sum_i \beta_i K(x_i, x_i) - \sum_{i,j} \beta_i \beta_j K(x_i, x_j)$$

Step 3: If $|f(x_i)| \leq \theta_1 R, \theta_1 \geq 1$, the value of θ_1 is a little greater than 1, then x_i is identified to be a fault sample point, and add it to T_{fl}.

If $|f(x_i)| \geq \theta_2 R, \theta_2 \geq \theta_1$,

the value of θ_1 is greater than 2, then x_i is identified to be a normal sample point, and add it to T_{nm}.

If $\theta_1 R \leq |f(x_i)| \leq \theta_2 R$,

substitute WKNN algorithm for clustering, pass parameters x_i, T_{sv}, T_{nm} and k to the WKNN, then x_i could be identify to be a normal point or a fault point.

Step 4: $T \leftarrow T - \{x_i\}$, return to step 1.

The WKNN algorithm in the algorithm takes a small part of the support vector obtained in the training stage as the representative point, and calculates the distance between the sample x_i to be identified and the representative point in the original space to find k neighbour of x_i. Since the support vector obtained after training only takes up a small part of the training set, it can avoid the problem of large computation in ordinary KNN algorithm.

When k-nearest neighbour method is used in feature space, the Euclidean distance of the two samples in the feature space is

$$D(x_i, x_j) = \|\varphi(x_i) - \varphi(x_j)\|$$

$$= (K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j))^{1/2} \quad (25)$$

But when the kernel function is gaussian kernel or exponential kernel, the distance formula can be simply adopted:

$$d(x_i, x_j) = \|x_i - x_j\| = ((x_i - x_j)^T (x_i - x_j))^{1/2} \quad (26)$$

Equation (20) refers to the Euclidean distance of the sample in the original input space, i.e. delete sample in the original input space. The following theorem guarantees that such a reduction is equivalent.

Theorem 1:

When the kernel function is gaussian kernel

$$K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$$

or exponential kernel,

$$K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|}{2\sigma^2})$$

the distance formula is equivalent by using equations (25) and (26).

Proof: We just need to prove that K nearest neighbours obtained by using these two distance formulas are identical.In fact, for both of these kernel functions:

$K(x_i, x_i) = K(x_j, x_j) = 1$, so

$$D(x_i, x_j) = (2 - 2\exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}))^{1/2} = (2 - 2\exp(-\frac{d(x_i, x_j)^2}{2\sigma^2}))^{1/2}$$

or

$$D(x_i, x_j) = (2 - 2\exp(-\frac{\|x_i - x_j\|}{2\sigma^2}))^{1/2} = (2 - 2\exp(-\frac{d(x_i, x_j)}{2\sigma^2}))^{1/2}$$

and

$D(x_i, x_j) \geq 0, d(x_i, x_j) \geq 0$, obviously, from the properties of power function and exponential function, we can get: For a fixed sample x_i , x_j , $D(x_i, x_j)$ is monotonically increasing with respect to $d(x_i, x_j)$. This indicates that all samples in the source input space and feature space are only different in compactness. The relative position of the sample does not change, and therefore the K neighbours of the sample are the same.

6. Results and Discussion

An example of circuit fault diagnosis was used to verify the feasibility and effectiveness of our fuzzy LSSVC-WKNN combination algorithm.

We select Gaussian Radial-basis function (RBF) as kernel function in the experiments. The expression of gaussian kernel function is

$$K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{\sigma^2}) \tag{27}$$

σ is the bandwidth in the Gaussian kernel. We set $\sigma = 0.6$ and $C = 10$

Experimental hardware environment is Intel Core i7 3.0GHz CPU and 8GB memory.

The analogue amplification circuit diagram is shown in figure 2, we use it as the simulation experiment to verify the fuzzy LSSVC-WKNN combination method.

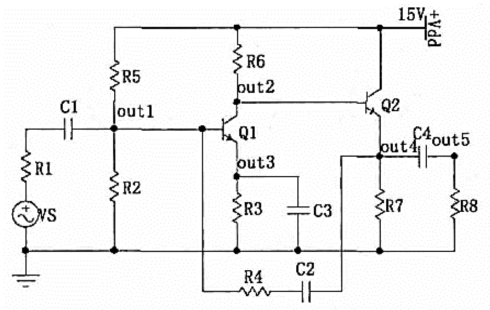

Figure 2. Analogue amplification circuit diagramme

In figure 2, Input signal source amplitude $V_m = 1000mV$, frequency $f = 1000Hz$, $R1 = 600\Omega$, $R2 = 1000\Omega$, $R3 = 500\Omega$, $R4 = R7 = R8 = 1200\Omega$, $R5 = 600\Omega$, $R6 = 1000\Omega$, four capacitors are equal, $C1 = C2 = C3 = C4 = 18\mu F$.

The fault type of analogue circuit is 7, that is, R2 open circuit, R1 short circuit, C1 short circuit, C2 short circuit, Q2 short circuit, Q1 open circuit and normal circuit. In order to obtain the sample data, voltage was obtained at the locations of out1, out2, out3, out4 and out5 in figure 2. A total of 200 samples were obtained, 100 were training sample data with fault type, and another 100 were test sample data.

The attributes obtained by adopting the method of ATS in section 3 are out1, out2 and out4. The feature vectors of some training samples optimized by ATS are shown in table 1.

Table 1. Reduced feature vectors

out1	out2	out4	Fault state
0	1	1	normal
0	1	1	R2 open circuit
0.5	0	0.5	R1 short circuit
0	1	0	Q2 short circuit
0	1	1	Q1 open circuit
0	0	0	C1 short circuit
0.5	0.5	0.5	C2 short circuit

At this point, the fuzzy LSSVC algorithm in section 4 is adopted to train the optimized feature vector of training samples, so as to obtain a trained clustering model. Then, the optimized test sample data is input into the fuzzy LSSVC-WKNN combination algorithm in section 5, its output result is used to identify it as fault sample or normal sample.

Table 2. Diagnosis accuracy comparison

Method	Correct samples	Fault samples	Accuracy(%)
1NN	94	6	94
C4.5	91	9	91
NaiveBayes	95	5	95
fuzzy LSSVC-WKNN	98	2	98

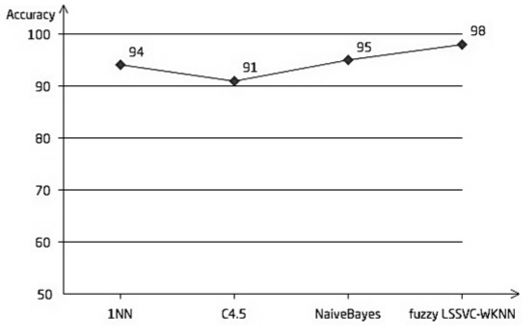

Figure 3. Diagnosis accuracy comparison diagram

It can be seen from table 2 and figure 3 that the method in this paper can correctly diagnose the fault of the circuit, and the fault diagnosis accuracy is as high as 98 %, 4 % higher than 1NN, 7 % higher than C4.5, and 3 % higher than NaiveBayes.

Table 3. The diagnosis time cost comparison

Method	Total number of samples	Time cost(ms)
1NN	100	2.6
C4.5	100	15.2
NaiveBayes	100	7.5
fuzzy LSSVC-WKNN	100	22.9

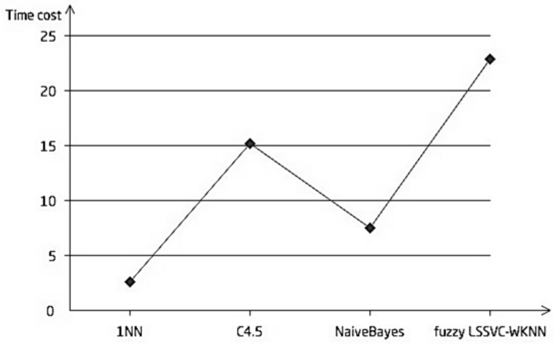

Figure 4. Diagnosis time cost comparison diagram

In table 3 and figure 4, the time cost of diagnosis is compared. The diagnosis time of 1NN algorithm is the shortest, only 2.6 ms, the second fastest is NaiveBayes, it's almost 5 ms slower than 1NN.

C4.5 is half as fast as NaiveBayes, our fuzzy LSSVC-WKNN takes 3 times longer than NaiveBayes.

Obviously, the simulation circuit experiment shows that the algorithm in this paper can accurately diagnose the fault, and because ATS is used for feature optimization, the fault diagnosis efficiency is high, which is a universal fault diagnosis method.

7. Conclusions

The drawback of the LSSVC method is the loss of sparseness. We apply fuzzy least squares to tackle this problem, a support degree is defined for each training point. the support degree corresponding to training data represents the data belonging to the degree of support vector. And we proposed a fuzzy LSSVC algorithm.

KNN algorithm has the category balance problem. In order to overcome it, we select the training samples with the same number of other categories from the category with the most obvious training data. KNN cannot identify the effect of attributes in dataset, so we give each of attributes a weight value according to the relationship between these attributes.

In order to further improve the diagnostic efficiency and accuracy of existing fault diagnosis algorithms, a method based on ATS to optimize feature vectors and adopt fuzzy LSSVC-WKNN combination algorithm for fault diagnosis is designed. Firstly, the fault diagnosis model is described, and then the feature vector is reduced by ATS method. The fuzzy LSSVC algorithm is adopted to train the optimized feature vector of training samples, so as to obtain a trained clustering model. Then, the fuzzy LSSVC-WKNN combination algorithm is used for fault diagnosis of test samples.

In the simulation circuit experiment, our method is compared against popular clustering algorithms, like 1NN, C4.5, and NaiveBayes. Experimental results show that our method can realize fault diagnosis of analogue circuit, and compared with other methods, the algorithm has higher diagnosis accuracy, is a feasible method to realize fault diagnosis.

8. Bibliographic References

[1] Ben-Hur A., Horn D., Siegelmann H. T., VAPNIK V., "Support vector clustering", in: *Machine Learning Research*, 2001, vol. 2, pp. 125-137.

[2] Vilen J., Pawel Z., Arkady B., "Ranking-based kernels in applied biomedical diagnostics using a support vector machine", in: *International Journal of Neural Systems*, 2011, vol. 21, no. 6, pp. 459-473.

[3] Yu C., Guangyu W., Fuqiang W., "Predicting financial distress of Chinese listed companies using rough set theory and support vector machine", in: *Asia-Pacific Journal of Operational Research*, 2011, vol. 28, no. 1, pp. 95-109.

[4] Widjaja D., Varon C., Dorado A., Suykens J., Van H., "Application of kernel principal component analysis for Single-lead-ECG-derived respiration", in: *IEEE Transactions on Biomedical Engineering*, 2012, vol. 59, no. 4, pp. 1169-1176.

[5] Suykens J. A. K., Debrabanter J., Lukas L., Vandewalle J., "Weighted least squares support vector machine: robustness and sparse approximation", in: *Neurocomputing*, 2002, vol. 48, pp. 85–105.

[6] Gavin C. C., Nicola L. C. T., "Improved sparse least-squares support vector machines", in: *Neurocomputing*, 2002, vol. 48, pp. 1025-1031.

[7] Fonseca E. S., Guido R. C., Scalassara P.R., et al, "Wavelet Time-frequency analysis and least squares support vector machines for the identification of voice disorders", in: *Computers in Biology and Medicine*, 2007, vol. 37, no. 4, pp. 571-578.

[8] Übeyli E. D., "Least squares support vector machine employing Model-based methods coefficients for analysis of EEG signals", in: *Expert Systems with Applications*, 2010, vol. 37, no. 1, pp. 233-239.

[9] Kruif B. J., Vries T. J., "Pruning error minimization in least squares support vector machines", in: *IEEE Transactions on Neural Networks*, 2003, vol. 14, no. 3, pp. 696–702.

[10] [10] KUH A., WILDE P., Comments on "Pruning error minimization in least squares support vector machines", in: IEEE *Transactions on Neural Networks*, 2007, vol. 18, no. 2, pp. 606–609.

[11] Jie Y., Abdesselam B., Son L. P., "A training algorithm for sparse LS-SVM using compressive sampling", in: *Proc. 35th Int. Acoustics Speech and Signal Processing Conf.*, 2010, pp. 2054-2057.

[12] Alshammari A., Zohdy M. A., Debnath D., Corser G., "Classification Approach for Intrusion Detection in Vehicle Systems", in: *Wireless Engineering and Technology*, 2018, vol. 9, pp. 79-94.

[13] Bista S., Chitrakar R, "DDoS Attack Detection Using Heuristics Clustering Algorithm and Naïve Bayes Classification", in: *Journal of Information Security*, 2018, vol. 9, pp. 33-44.

[14] Hosseini L., Kandovan R. S., "Hyperspectral Image Classification Based on Hierarchical SVM Algorithm for Improving Overall Accuracy", in: *Advances in Remote Sensing*, 2017, vol. 6, pp. 66-75.

[15] Liu G. Q., Jiang M. R., Cun H. L., Shi Z. Z., Hao J. Y.,

[16] "An Automatic Text Region Positioning Method for the Low-Contrast Image", in: *Journal of Computer and Communications*, 2017, vol. 5, pp. 36-49.

[17] Mack J., Trakowski A., Rist F., Herzog K., Töpfer R., Steinhage V, "Experimental Evaluation of the Performance of Local Shape Descriptors for the Classification of 3D Data in Precision Farming", in: *Journal of Computer and Communications*, 2017, vol. 5, pp. 1-12.

[18] Sakiz F., Sen S, "A Survey of Attacks and Detection Mechanisms on Intelligent Transportation Systems: VANETs and IoV", in: *Ad hoc Networks*, 2017, vol. 61, pp. 33-50.

[19] Sheng S., Chuiwei L., Guojun Z., "A Fuzzy Least Squares Support Vector Clustering Algorithm", In: *Electrical Electrotehnica, Electronica, Automatica (EEA)*, 2016, vol. 64, no. 2, pp. 119-123.

[20] Sheng S., Chuiwei L., Guojun Z., "Fault Diagnosis of Distribution Systems Based on Least Squares Support Vector Clustering Algorithm", In: *Electrotehnica, Electronica, Automatica (EEA)*, 2017, vol. 65, no. 2, pp. 133-139.

[21] Ruban N., Razzhivin I., Suvorov A., Askarov A., Kievets A.,"Research of influence of renewable energy sourcesand FACTS on the out-of-step protection operation", in: *Electrotehnica, Electronica, Automatica (EEA)*, 2019, vol. 67, no. 3, pp. 14-21.

[22] Ameur I., Benmiloud M., Benalia A.,"A Novel Hybrid Control for Maximizing Power in a Photovoltaic System", in: *Electrotehnica, Electronica, Automatica (EEA)*, 2019, vol. 67, no. 3, pp. 5-13.

[23] Sivaraj C., Phonse P. J. A., Janakiraman T. N., "Energy-efficient and Load Distributed Clustering Algorithm for Dense Wireless Sensor Networks", in: *International Journal of Intelligent Systems and*

Applications, 2017, vol. 9, no. 5, pp. 34-42.

[24] Zhengbing H., Yevgeniy V. B., Oleksii K. T., Viktoriia O. S., "Possibilistic Fuzzy Clustering for Categorical Data Arrays Based on Frequency Prototypes and Dissimilarity Measures". In: *International Journal of Intelligent Systems and Applications*. 2017, vol. 9, no. 5, pp. 55-61.

[25] Thomas L. ,Christopher T., "Microarray Gene Retrieval System Based on LFDA and SVM", in: *International Journal of Intelligent Systems and Applications*, 2018, vol. 10, no. 1, pp. 9-15.

Funding Sources

It is a pleasure to acknowledge the contributions of Prof. Chuiwei Lu and Dr. Guojun Zhang, both of whom participated in a number of fruitful discussions. This work was financially supported by the Research Project of Education Department of Hubei Province (No.B2017256), Science and Technology Innovation Team of Outstanding Youth of Hubei Polytechnic University (No.13xtz06).

Authors' Biographies

Sheng Sun was born in Huang gang (China), on January 9, 1978.
He graduated from the Huazhong University of Science and Technology, Computer Science in Wuhan (China), in 1999.
He received the PhD degree in Computer Science from Huazhong University of Science and Technology (Wuhan, China) in 2011.
His present research interests concern: computer security and reliability, computer communications.
His present research interests concern: software engineering and data mining.
e-mail: 1106049065@qq.com

Chuiwei Lu was born in Huangshi (China), on March 18 , 1973.
He graduated from the Northwestern polytechnical university, Institute of electrical and mechanical in Xian (China), in 1995.
He received the PhD degree in Computer Science from Huazhong University of Science and Technology (Wuhan, China) in 2010.
He is professor at the Hubei Polytechnic University, in Huangshi (China).

Guojun Zhang was born in Huangshi (China), on November 9, 1973.
He graduated from the Central China Normal University, School of Computer in Wuhan (China), in 1992.
He received the PhD degree in Computer Science from Huazhong University of Science and Technology (Wuhan, China) in 2010.
He is associate professor at the Hubei Polytechnic University, in Huangshi (China).
His present research interests concern: software engineering and data mining.
e-mail: 1106049065@qq.com