

Article

Developing Support Vector Machine with New Fuzzy Selection for the Infringement of a Patent Rights Problem

Chih-Yao Chang ^{1,2} and Kuo-Ping Lin ^{3,4,*}

¹ Graduate Institute of Technology, Innovation & Intellectual Property Management, National Cheng Chi University, Taipei 116302, Taiwan; yao@tedr.org.tw

² Taiwan Development & Research Academia of Economic & Technology, Taipei 104, Taiwan

³ Department of Industrial Engineering and Enterprise Information, Tunghai University, Taichung 40704, Taiwan

⁴ Faculty of Finance and Banking, Ton Duc Thang University, Ho Chi Minh City 758307, Vietnam

* Correspondence: kplin@thu.edu.tw; Tel.: +886-4-2359-4319 (ext. 111)

Received: 16 June 2020; Accepted: 30 July 2020; Published: 1 August 2020

Abstract: Classification problems are very important issues in real enterprises. In the patent infringement issue, accurate classification could help enterprises to understand court decisions to avoid patent infringement. However, the general classification method does not perform well in the patent infringement problem because there are too many complex variables. Therefore, this study attempts to develop a classification method, the support vector machine with new fuzzy selection (SVMFS), to judge the infringement of patent rights. The raw data are divided into training and testing sets. However, the data quality of the training set is not easy to evaluate. Effective data quality management requires a structural core that can support data operations. This study adopts new fuzzy selection based on membership values, which are generated from fuzzy c-means clustering, to select appropriate data to enhance the classification performance of the support vector machine (SVM). An empirical example based on the SVMFS shows that the proposed SVMFS can obtain a superior accuracy rate. Moreover, the new fuzzy selection also verifies that it can effectively select the training dataset.

Keywords: classification; patent infringement; support vector machine; fuzzy selection

1. Introduction

Nowadays classification models have various limitations related to the use of a single model. Data preprocessing plays a significant role in the entire dataset. In the data preprocessing, detecting outliers which appear to not belong in the data is one of the important methods, and can be caused by human error, such as mislabeling, transposing numerals, and programming bugs. Outliers corrupt the results to a small or large degree, depending on the circumstances, if they are not removed from the raw dataset. This study develops a fuzzy selection strategy that addresses fuzzy membership for selecting data to be eliminated from datasets. New algorithms should provide high quality and clean data to treat the noise (smart data) in Big Data analysis problems [1]. Therefore, this study attempts to handle noisy data by proposing a new algorithm for unstructured datasets. Some studies have developed outlier detection methods, for example, van der Gaag [2] used FDSTools noise profiles to obtain training datasets and a test set to analyze the impact of FDSTools noise correction for different analysis thresholds. This method was able to obtain a higher quality training dataset, leading to improved performance. Niu and Wang [3] proposed a combined model to achieve accurate prediction results. The combined model included complete empirical mode decomposition ensemble, four neural network models, and a linear model. Cai et al. [4] adopted Kalman filter-deduced noisy datasets. The Kalman filter is insensitive to non-Gaussian noises because it uses the maximum

correntropy criteria. Numerical experiments have shown to outperform this model on four benchmark datasets for traffic flow forecasting. Liu and Chen [5] conducted a comprehensive review of data processing strategies in wind energy forecasting models. This research mentioned that the existing data-driven forecasting models attach great significance to the proper application of data processing methods. Ma et al. [6] developed an unscented Kalman filter (UKF) with the generalized correntropy loss (GCL) which can be termed GCL-UKF. GCL-UKF has been used to estimate and forecast the power system state. Numerical simulation results have validated the efficacy of the proposed methods for state estimation using various types of measurement. Wang et al. [7] developed wavelet de-noising (WD) and Rank-Set Pair Analysis (RSPA), which is a hybrid model. RSPA takes full advantage of a combination of the two approaches to improve forecasts of hydro-meteorological time series. Florez-Lozano et al. [8] developed an intelligent system that combined both classic aggregation operators and neural and fuzzy systems. These studies proposed pre-processing methods to effectively handle datasets for various forecasting methods. Therefore, this study develops a fuzzy selection strategy that addresses fuzzy membership. The new fuzzy selection operator based on the fuzzy clustering algorithm is carefully formulated to ensure that there are better members of the dataset in the proposed classification system.

This study develops the support vector machine (SVM) classification model with new fuzzy selection to improve the performance of the classification problem. Supervised classification is the essential technique used for extracting quantitative information from the database, and the SVM is one a popular classifier. Tang et al. [9] developed a joint segmentation and classification framework for sentence-level sentiment classification. Their method simultaneously generates useful segmentations and predicts sentence-level polarity based on the segmentation results. The effectiveness of the approach was verified by applying it to sentiment classification. Jiang et al. [10] proposed a method for the effective classification and localization of mixed sources. The advantage of this method is that it could make good use of known information in order to distinguish the distances of sources from mixed sources and estimate the range parameters of near-field sources. Kasabov et al. [11] developed a new and efficient neuromorphic approach to the most complex rich spatiotemporal brain data (STBD) and functional magnetic resonance imaging (fMRI) data. Shao et al. [12] developed a prototype-based classification model which evolving data streams. Building upon the techniques of error-driven representativeness learning, P-Tree based data maintenance, and concept drift handling, SyncStream allows dynamic modeling of the evolving concepts and supports good prediction performance. Wang et al. [13] developed Noise-resistant Statistical Traffic Classification (NSTC) to solve the traffic classification problem. NSTC could reduce the noise and reliability, thereby improving the classification performance. Phan et al. [14] developed a joint classification-and-prediction framework based on convolutional neural networks (CNNs) for automatic sleep staging. The CNNs divide the dataset into nontransition and transition sets and explored = how different frameworks perform on them. Basha et al. [15] proposed the interval principal component analysis to detect faults in the Tennessee Eastman (TE) Process with a higher degree of accuracy than other methods. The interval principal component analysis method has been capable of maintaining a high performance rate even at low GLR window sample sizes and low interval aggregation window sizes. The main focus of these studies was to solve actual problems using new classification methods. Moreover, pre-processing methods, such as those presented in [12] and [13], could improve performance. Therefore, this study develops a fuzzy selection strategy that addresses how fuzzy membership and the support vector machine (SVM) method provide an effective way to perform supervised classification. The recent SVM literature is summarized and shown in Table 1. It can be observed that the UCI dataset [16] has been examined in many studies, and many studies have developed hybrid SVMs for improving the classification performance.

Table 1. Summary of studies of the support vector machine (SVM) methods conducted since 2016.

Author(s)	Year	Methods	Applied Fields
Ekong et al. [17]	2016	Interval type-2 fuzzy SVM	Epileptic seizure phases
Shen et al. [18]	2016	SVM with fruit fly optimization algorithm	Medical diagnosis
Wang et al. [19]	2016	Locally linear SVMs	UCI dataset
Zhong et al. [20]	2016	SVM with novel Bayesian inference	Industrial survey data
Qi et al. [21]	2016	New Deep SVM	UCI dataset
Liu et al. [22]	2017	SVM	One-vs-one (OVO) strategy
Zhang et al. [23]	2017	Multiple birth SVM	UCI dataset
Utkin and Zhuk [24]	2017	One-class classification SVM	UCI dataset
Gonzalez-Abril et al. [25]	2017	Modified SVM	UCI dataset
Kusakci et al. [26]	2017	Competitive SVM Teams	Autonomous human chromosome
Richhariy and Tanveer [27]	2018	Universum support vector machine	Electroencephalogram
Ougiaroglou et al. [28]	2018	SVM	KEEL-dataset repository
de Lima et al. [29]	2018	Least squares twin multi-class SVM	UCI dataset
Tang et al. [30]	2019	Regular simplex SVM	Benchmark datasets
Yang and Dong [31]	2019	SVM with generalized quantile loss	UCI dataset
Okwuashi and Ndehedehe [32]	2020	Deep SVM	Hyperspectral image

Furthermore, some literature has combined the fuzzy *c*-means (FCM) with SVM for improving the performance of the classifier ([33–38]). These studies used the clustering label of FCM as the preprocess mechanism for improving the SVM classifier. This study adopts the roulette wheel selection with a membership function of FCM to select appropriate data for the training set. Traditional roulette wheel selection is based on probability to select possible eliminates in genetic algorithms. This study adopts membership values of FCM to roulette wheel selection for the possible eliminates of training datasets.

The purpose of this study is to develop a new classification method that combines SVM with new fuzzy selection. A core component of this study is the development of a new fuzzy selection method, roulette wheel selection with a membership function, to select appropriate data for the training set. The proposed methodology draws on the advantages of fuzzy clustering, roulette wheel selection, and SVM to effectively handle the dataset, reduce outlier data, and improve classification performance. The remainder of this paper is organized as follows: Section 2 presents the proposed support vector machine with new fuzzy selection method and also introduces the new fuzzy selection method; Section 3 provides the research design of the support vector machine with new fuzzy selection (SVMFS) for the actual infringement of patent rights problem; and Section 4 offers conclusions and suggestions for further research.

2. Support Vector Machine with New Fuzzy Selection

Generally, the raw data are divided into training and testing sets. However, the data quality of training set is not easily evaluated. Effective data quality management requires a structural core that can support data operations. The data quality of training set is a very important issue in classification problem. This study attempts to adopt the new fuzzy selection method to select appropriate data for enhancing the classification performance of the SVM. Figure 1 shows the flowchart of the proposed SVMFS method.

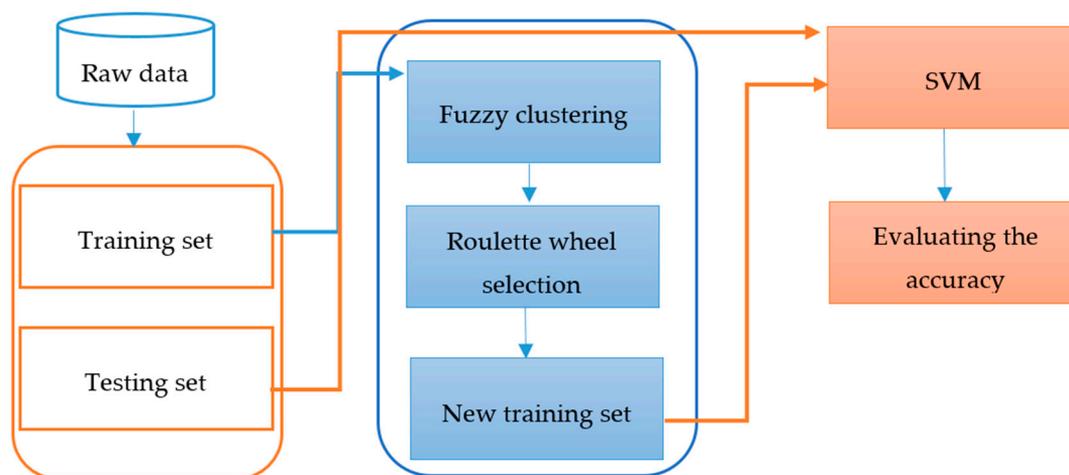


Figure 1. An illustration of the proposed support vector machine with new fuzzy selection (SVMFS) method.

2.1. New Fuzzy Selection

First, the dataset uses the fuzzy clustering algorithm to determine fuzzy membership. In this study, we adopt the fuzzy c-means algorithm ([39–41]). The distance function can be defined as a membership function of clustering. Therefore, the closeness of the data to the multi-cluster center can be expressed by the degree of membership of the distance function. The FCM objective function from [39] can be formulated as:

$$\sum_{j=1}^c \sum_{i=1}^N (\mu_{ij}^{(l)})^\omega \times d_{ij}(\mathbf{x}_i, \mathbf{o}_j^{(l)}) \tag{1}$$

where c is the number of clusters to classify, ω is a parameter for updating the clustering membership functions, l is number of epochs to execute. In FCM, the initial memberships of the data $x_i, i = 1, \dots, N$ with crisp input–output to the clusters $j (j = 1, \dots, c)$ are generated randomly. The initial l value is set to 0. FCM optimizes the objective function by continuously updating the membership functions and centers of clusters until the difference in updating the membership functions is smaller than the tolerance for the solution accuracy. d is the distance between data x_i and the center of cluster j, O_j . The equations used to update membership functions and centers of clusters can be expressed as follows:

$$\begin{aligned} d_{ij}(\mathbf{x}_i, \mathbf{o}_j^{(l)}) &= (\mathbf{x}_i - \mathbf{o}_j^{(l)})^T I(\mathbf{x}_i - \mathbf{o}_j^{(l)}) \\ \mu_{ij}^{(l)} &\in [0, 1], \sum_{j=1}^c \mu_{ij}^{(l)} = 1, \text{ and } 0 < \sum_{i=1}^N \mu_{ij}^{(l)} < N. \end{aligned} \tag{2}$$

$$\begin{aligned} \mu_{ij}^{(l+1)} &= \begin{cases} \left(\frac{d(\mathbf{x}_i, \mathbf{o}_j^{(l+1)})}{\sum_{h=1}^c \left(\frac{d(\mathbf{x}_i, \mathbf{o}_h^{(l+1)})}{d(\mathbf{x}_i, \mathbf{o}_h^{(l+1)})} \right)^{(1/(\omega-1))} } \right)^{-1}, & \text{if } d(\mathbf{x}_i, \mathbf{o}_j^{(l+1)}) > 0, \\ 1, & \text{if } d(\mathbf{x}_i, \mathbf{o}_j^{(l+1)}) = 0, \end{cases} \\ \mathbf{o}_j^{(l+1)} &= \frac{\sum_{i=1}^N \mu_{ij}^{(l+1)} \mathbf{x}_i}{\sum_{i=1}^N \mu_{ij}^{(l+1)}}. \end{aligned} \tag{3}$$

Traditional roulette wheel selection is similar to a roulette wheel in a casino which is assigned to each of the possible selections based on probability. This study develops the new fuzzy roulette wheel

selection is based on fuzzy membership function of FCM. Based on the final membership function, the fuzzy selection strategy determines which of the members in the current $U^{(l)}$ will have an even higher membership. The selection operator is carefully formulated to ensure that better members of the dataset have a greater probability of being selected to improve the accuracy of classification. The procedures of selection for worse members of the dataset still could be selected which are small probability, and this is important to ensure that the noise filtering process is reasonable. Fitter datasets are more likely to be selected as training (X_{train}) and testing (X_{test}) datasets. The Algorithm 1 of the new fuzzy roulette wheel selection is displayed as following:

Algorithm 1: Fuzzy Roulette Wheel Selection with Membership Function

```

For  $I < \text{Max size}$ 
  Generate Max size random number  $\gamma$ 
  Calculate cumulative membership  $U$ , total fitness  $\mu_i$  and sum of proportional membership  $\Sigma u_i$ 
  Spin the wheel Max size times
  If  $\Sigma u_i < \gamma$  then
    Select the dataset as training set ( $X_{train}$ ) and testing ( $X_{test}$ ) datasets, otherwise, index to outlier dataset
  End if
End
Return dataset with membership value proportional to the size of selected wheel section
End Procedure
    
```

2.2. SVMFS

The support vector machine (SVM) proposed by Vapnik [42] uses classification techniques drawn from statistical learning theory ([42,43]). The SVM performs a binary classification using the optimal separation hyperplane, which is nonlinearly mapped into the high-dimensional feature space. The data are linearly separated using SVM, which trains the linear machine to obtain the best hyperplane, and the decision function is estimated through the nonlinear class boundary, which is based on the support vector. The data are sorted according to the maximum distance between hyperplanes or the nearest training point. The training points closest to the optimal separating hyperplane are called support vectors:

$$K(Xt, z) = \sum_{i=1}^{n_H} \lambda_i \phi_i(Xt_i) \phi_i(z), \tag{4}$$

where $x, z \in \mathcal{R}^n$, and n_H are the dimensions of H . Equation (4) depicts Mercer’s required conditions for any square integrable function $g(x)$. The kernel function can be expressed as the inner product. Therefore, having a positive semi-definite kernel implies that kernel K is separately used to solve integral equations. Several choices are possible for kernel $K(\cdot)$. The kernel function enables the SVM to function within very large dimensional feature spaces without making explicit computations in the space. Computations are completed in another space subsequent to the application of this kernel trick. One begins with a formulation in the primal weight space with a high dimensional feature space through the application of the transformation function $\phi(\cdot)$. The problem cannot be solved in the primal weight space, but it can be handled in the dual space by applying the kernel functions. In this way, we are able to implicitly compute the problem within a high dimensional feature space. The extension from linear SVM classifiers to nonlinear SVM classifiers can be completed by replacing x with $\phi(x)$ and applying the kernel method wherever possible:

$$\int K(x, z)g(x)g(z)dx dz \geq 0 \tag{5}$$

$$K(x, z) = \phi(x)^T \phi(z) \tag{6}$$

Originally, SVM was designed for two-class classification. Based on the process of determining the separate boundary and the maximum distance to the closest points, SVM derives a class choice,

called support vectors (SVs), for the training dataset. SVM can avoid a potential misclassification in the testing data by minimizing the structural risk rather than the empirical risk. Therefore, the SVM classifier demonstrates a better generalization performance than that of other traditional classifiers. First, we give a training dataset $D = \{X_{train}, Y_{train}\}$, where $X_{train} \in \mathfrak{R}^n$ is the input vector which is selected from fuzzy roulette wheel selection with a known binary output label, $Y_{train} \in \{0, 1\}$. Then, the classification function is specified by:

$$y_i = f(x_i) = w^T \phi(x_i) + b \tag{7}$$

where $\phi: \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ is the feature mapping of the input space to a high dimensional feature space. The data points become linearly separable by a hyperplane defined by the pair $(w \in \mathfrak{R}^m, b \in \mathfrak{R})$ [35]. The optimal hyperplane separating the data is expressed as Equation (8):

$$\begin{aligned} \text{Minimize} \quad & \Phi(w) = \|w\|^2 / 2 \\ \text{Subject to} \quad & y_i [w^T \phi(x_i) + b] \geq 1 \quad i = 1, \dots, N \end{aligned} \tag{8}$$

where $\|w\|$ is the norm for a normal weights vector of a hyperplane. This constrained optimization problem is solved by the following primal Lagrangian form:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i [y_i (w^T \phi(x_i) + b) - 1] \tag{9}$$

where α_i are the Lagrange multipliers between 0 and C. Applying the Karush–Kuhn–Tucker conditions, solutions for the dual Lagrangian problem, α_i^0 , decide the parameters w_0 and b_0 of the optimal hyperplane.. Next, the decision function is generated by Equation (10):

$$d(x_i) = \text{sign}(w_0^T \phi(x_i) + b_0) = \text{sign}\left(\sum_{i=1}^N \alpha_i^0 Y_i K(x, x_i) + b_0\right), \quad i = 1, \dots, N \tag{10}$$

$K(x, x_i)$ is the kernel function and should satisfy Mercer’s condition, as mentioned previously. In addition, the value of the kernel function is equal to the inner product of two vectors, x and x_i , in the feature space $\phi(x)$ and $\phi(x_i)$. Note that in the case of the radial basis function (RBF) kernels, which are represented in Equation (11), one has only two additional tuning parameters (C, σ):

$$K(x, x_i) = \exp(-\|x - x_i\|_2^2 / \sigma^2) \tag{11}$$

In addition, having suitable parameters will effectively improve the performance of the SVM.

3. Patent Infringement Problem

A patent infringement problem was executed with various classification methods using the standard SVM, principal component analysis (PCA) [44] + SVM, least square support vector machine (LSSVM), back propagation neural network (BPNN), least square support vector machine with new fuzzy selection (LSSVMFS), and the proposed SVMFS. The definition of patent infringement refers to the unlawful conduct of a patented invention without the permission of the patent holder. Moreover, the definition of patent infringement may vary by jurisdiction, but it typically includes using or selling the patented invention. The patent infringement applications can adopt the user modelling technique. Researchers use the classification model to model court decisions according to collected datasets. In this study, the patent infringement dataset came from Taiwan courts. This study attempts to improve the accuracy of the patent infringement dataset in classification fields.

3.1. Data

The data used in this investigation were gathered from law offices. This study develops the SVMFS to the real patent infringement classification problem which could assist lawyer judging patent infringement cases by data-driven technique. In total, 93 original data points were divided into training and testing datasets. The distributions of the data are shown in Table 2. The key/input variables include “Months”(M), “International Patent classification—Main item”(IPC-M), “International Patent classification—Subitem”(IPC-S), “Case category”(CC), “Case category—Main item”(CC-M), “Case category—Subitem”(CC-S), “Service fee”(SF), “Client”(CL), “Authorized capital”(AC), “Agent”(A), and “Expected judgment”(EJ), and the decision variable is “Court decisions”(CD), as shown in Table 2. The numerical expression is defined by lawyer’s options. Table 2 also gives a summary of the statistical analyses conducted, such as the maximum, minimum, average, and standard deviation values for various variables. Table 2 shows that the variation in the authorized capital is large, which may affect the classification performance. The court decision was that a patent infringement (69 cases) had occurred in about 74.2% of 93 cases. Table 3 shows the input data example of patent infringement cases for various classification models. Moreover, the normalization is employed for reducing the effect of numbers for categories in this study. Table 4 displays the correlation matrix of the variables. The correlations of all variables could be observed are much lower and are not statistically significant in the patent infringement example.

Table 2. Summary of statistics and variable information of patent infringement cases.

Variables	Descriptions	Numerical Expression	Maximum	Minimum	Average	Standard Deviation
M	The month of the case	1 to 12: means January to December	12	1	6.602	3.140
IPC-M	Uniform classification of patent documents in various countries	0 to 8 categories	8	0	4.839	2.705
IPC-S	Uniform classification of patent documents in various countries	1 to 67 categories	67	1	19.989	24.078
CC	Three types of patent	1: Invention patent 2: Utility model patent 3: Design patent	3	1	1.312	0.724
CC-M	Number of rights in patents	1: Means only one right in patents.	28	1	3.108	4.390
CC-S	The number of items in a patent that can be independently claimed	1: Means only one independently claimed.	4	1	1.086	0.483
SF	The fees charged for each case.	1 to 14 represent different service charges.	14	1	7.742	3.077
CL	The identity of the client	1: Public company 2: Medium-sized and small companies 3: Person 4: Court 5: Others	5	1	3.269	1.275
AC	Demonstration of the scale of resources	Amount of authorized capital	245,000,000,000	0	6,234,745,161	43,447,535,187
A	Law office	1: Small law office 2: Medium-sized law office 3: Larger law office 4: Famous law office	4	1	2.129	0.816
EJ	Expected judgment from client	1: Infringement	1	0	0.441	0.500

0: Non- infringement						
Decision variables						
CD	Final judgment from court.	1: Infringement	1	0	0.742	0.438
0: Non-infringement						

Table 3. Input data example of patent infringement cases

	M	I	IPC-M	IPC-S	CC	CC-M	CC-S	SF	CL	AC	A	EC	CD
Case 1	1	7	1	3	5	1	11	3	25,000,000	2	1	1	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
Case 93	12	5	4	1	2	1	0	5	0	1	3	0	

Table 4. The correlation matrix of the variables

		M	IPC-M	IPC-S	CC	CC-M	CC-S	SF	CL	AC	A	EC	CD
M	Pearson Correlation	-0.142	0.154	-0.163	-0.153	-0.060	-0.227	-0.158	-0.126	-0.088	0.118	-0.136	
	Sig. (2-tailed)	0.171	0.138	0.116	0.140	0.566	0.028	0.129	0.226	0.401	0.258	0.193	
IPC-M	Pearson Correlation		-0.533	0.200	0.394	0.325	0.280	0.148	0.227	-0.065	0.197	-0.028	
	Sig. (2-tailed)		0.000	0.053	0.000	0.001	0.006	0.156	0.028	0.537	0.057	0.789	
IPC-S	Pearson Correlation			-0.209	-0.191	-0.066	-0.229	-0.076	0.043	0.093	0.020	0.120	
	Sig. (2-tailed)			0.043	0.066	0.525	0.026	0.466	0.683	0.370	0.851	0.251	
CC	Pearson Correlation				0.004	-0.108	0.270	-0.227	-0.164	-0.067	-0.058	0.100	
	Sig. (2-tailed)				0.971	0.301	0.009	0.028	0.115	0.523	0.576	0.338	
CC-M	Pearson Correlation					0.632	-0.034	0.102	0.512	0.212	0.181	0.142	
	Sig. (2-tailed)					0.000	0.745	0.328	0.000	0.040	0.081	0.171	
CC-S	Pearson Correlation						-0.006	0.091	0.483	0.267	0.260	-0.019	
	Sig. (2-tailed)						0.954	0.382	0.000	0.009	0.011	0.858	
SF	Pearson Correlation							-0.138	-0.077	0.006	-0.054	0.016	
	Sig. (2-tailed)							0.185	0.460	0.957	0.603	0.878	
CL	Pearson Correlation								0.229	-0.293	0.276	-0.016	
	Sig. (2-tailed)								0.026	0.004	0.007	0.876	
AC	Pearson Correlation									0.067	0.201	0.106	
	Sig. (2-tailed)									0.520	0.052	0.310	
A	Pearson Correlation										0.099	-0.001	
	Sig. (2-tailed)										0.344	0.990	
EC	Pearson Correlation											-0.107	
	Sig. (2-tailed)											0.306	

3.2. Statistical Comparison with Different Numbers of Testing Datasets

This research coded the classification models using MATLAB R2019b (The MathWorks, Inc. Natick, MA, USA). The classification models can be divided into new fuzzy selection and least square support vector machine (LSSVM)/SVM models. The various numbers of clusters used to classify values of FCM ($c = 3, 4, \text{ and } 5$) were tested with X_{test} numbers of 10, 20, and 30 in the least square support vector machine with fuzzy selection (LSSVMFS) and SVMFS models. Figure 2 shows the training error of fuzzy clustering in patent infringement datasets. The experimental results show that the training error of fuzzy clustering with different number of clusters converges with excellent stability. Two measuring indices employed are the true positive rate (TPR) and the accuracy. Then, the measuring indexes is specified by:

$$TPR = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \tag{12}$$

$$Accuracy = \frac{\text{True positive} + \text{True negative}}{\text{Total population}} \tag{13}$$

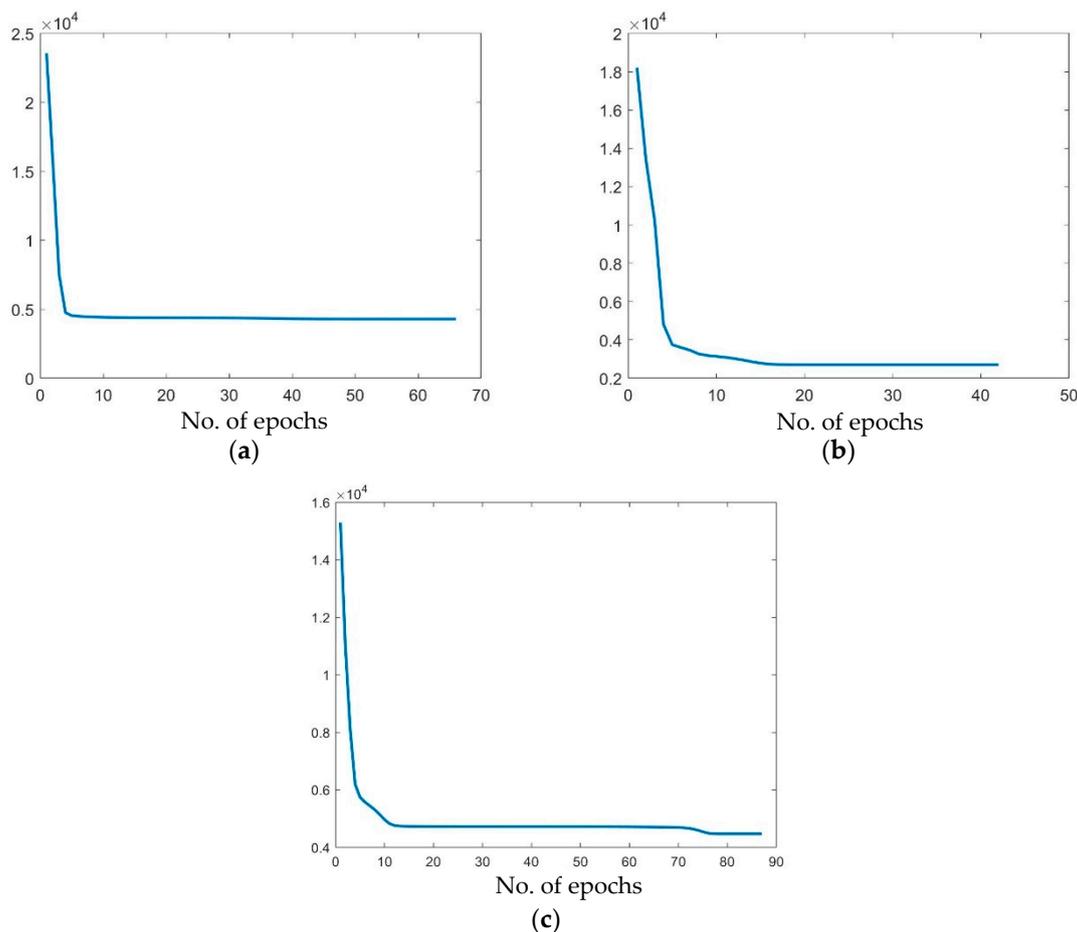


Figure 2. Illustration of the training error of fuzzy clustering in patent infringement datasets. (a) $c = 3$; (b) $c = 4$; (c) $c = 5$.

In order to identify LSSVM/SVMFS model parameter (Number of c) and obtain average and variance performances, LSSVM/SVMFS execute 20 times each $c = 3, 4, \text{ and } 5$. Table 3 indicates that the SVMFS model with $c = 3, 4, \text{ and } 5$ in X_{test} numbers of 10, 20, and 30 can has a higher accuracy than the LSSVMFS. Furthermore, the SVMFS with $c = 3$ outperforms (average and variance are 0.870 and

0.103, respectively) in X_{test} numbers of 10. The SVMFS with $c = 5$ outperforms (the average and variance are 0.875 and 0.06, respectively) in X_{test} numbers of 20. The SVMFS with $c = 4$ outperforms (the average and variance are 0.85 and 0.045, respectively) in X_{test} numbers of 30. All the best performance in different X_{test} numbers are bolded in Table 3. Figures 3–5 illustrate boxplots of different numbers of testing datasets for the LSSVRFS/SVRFS model with various cluster values. In the box plots of Figures 3–5, the lowest point is the minimum of the accuracy and the highest point is the maximum of the accuracy with different numbers of testing datasets. As demonstrated in the sensitivity analysis provided in Table 5 and Figures 3–5, the SVRFS model has smaller variance with the *cluster* value in different numbers of testing datasets. Therefore, with various *cluster* values, the SVRFS model can provide stable results with various numbers of X_{test} and give an accurate classification.

Table 5. Summary of statistics and variable information of patent infringement datasets.

No. of X_{test}	10			20			30		
No. of c	3	4	5	3	4	5	3	4	5
LSSVM	0.805	0.785	0.815	0.7725	0.800	0.7825	0.7405	0.7683	0.78
FS	(0.105)	(0.134)	(0.087)	(0.100)	(0.06)	(0.092)	(0.06)	(0.047)	(0.057)
SVMFS	0.870	0.865	0.85	0.8625	0.873	0.875	0.830	0.85	0.846
	(0.103)	(0.098)	(0.105)	(0.062)	(0.079)	(0.06)	(0.05)	(0.045)	(0.06)

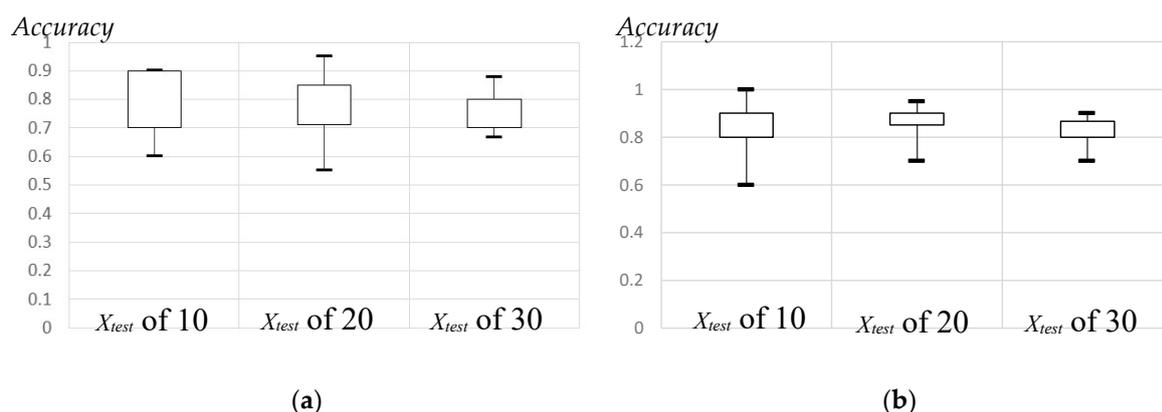


Figure 3. Boxplots of the proposed method with an $c = 3$. (a) LSSVMFS; (b) SVMFS.

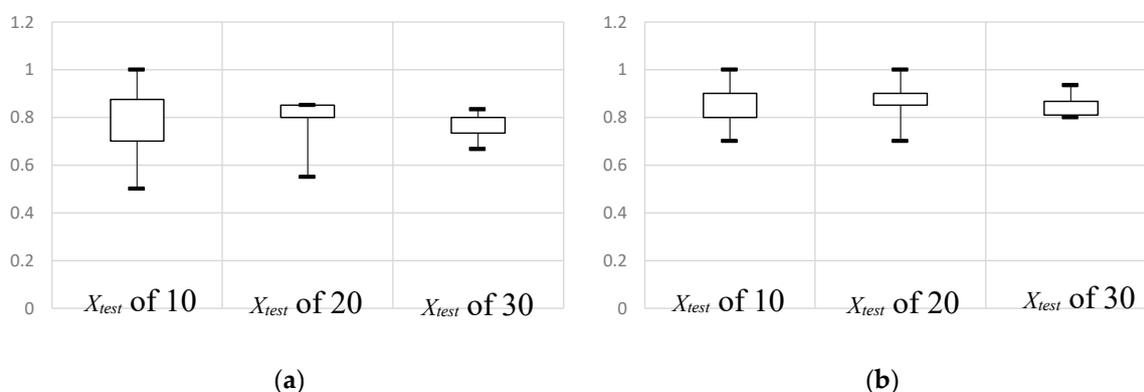


Figure 4. Boxplots of the proposed method with $c = 4$. (a) LSSVMFS; (b) SVMFS.

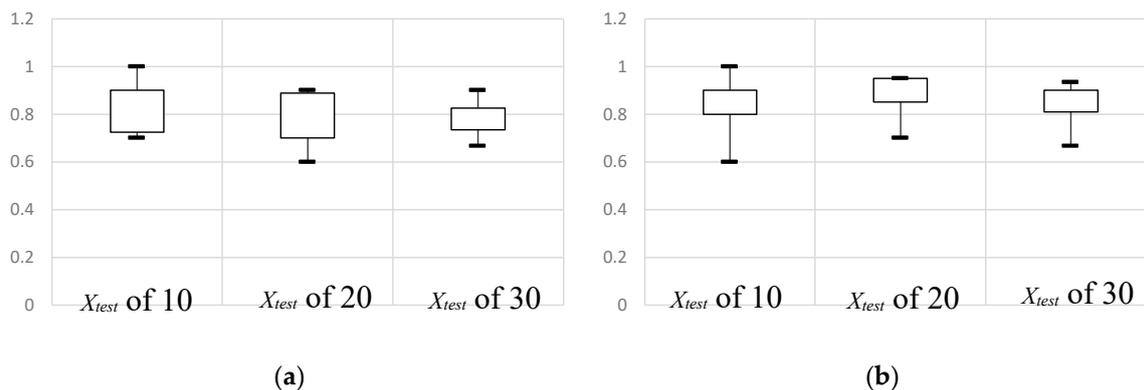


Figure 5. Boxplots of the proposed method with $c = 5$. (a) LSSVMFS; (b) SVMFS

3.3. The Results Analysis

Table 6 shows the results of the patent infringement datasets using various classification methods. SVMFS obtained higher accuracy rates of 0.87, 0.86, and 0.83 and TPN of 1, 1, and 1 with 10, 20, and 30 testing datasets, respectively, than the traditional SVM, principal component analysis PCA + SVM, LSSVM, and back propagation neural network (BPNN) methods. The PCA could be observed that has not improved the accuracy of SVM in the patent infringement example. The reason may be the correlation of the variables are much lower. Hence the reducing variables of PCA in the patent infringement example may not obviously enhance SVM classification. Our proposed new fuzzy selection algorithm obviously improves the SVM/LSSVM classification performance in patent infringement datasets. The SVMFS also can effectively discover the structure of patent infringement datasets. The reason may be the membership of FCM method is obtain more information of training set for uncertain classification problem.

From observing these experiments, this study can conclude (1) that the new fuzzy selection and SVM method can improve the classification performance and (2) that quality of the data for processing is improved by new fuzzy selection method when the structure of datasets is very complex.

Table 6. Classification results for patent infringement datasets.

Methods		No. of Testing set		
		10	20	30
SVM	TPN	1	1	1
	Accuracy	0.80	0.70	0.73
PCA + SVM	TPN	1	1	1
	Accuracy	0.80	0.70	0.73
LSSVM	TPN	1	0.93	0.91
	Accuracy	0.60	0.75	0.70
BPNN	TPN	0.88	0.86	0.81
	Accuracy	0.80	0.65	0.7
SVMFS	TPN	1	1	1
	Accuracy	1	0.95	0.93
LSSVMFS	TPN	1	1	0.96
	Accuracy	0.9	0.9	0.87

4. Conclusions

This study firstly developed a SVMFS and examined it using patent infringement datasets. The results indicate that SVMFS offers a promising alternative for classification. Overall, the SVMFS can provide a more stable and better performance with a higher level of accuracy. The superior performance of SVMFS can be attributed to several factors, as follows: (1) the new fuzzy selection can

enhance the quality of the data for processing. In the fuzzy cluster algorithm, the closeness of the data to the multi-cluster center can be expressed by the degree of membership of the distance function. Based on the degree of membership of the distance function, the roulette wheel selection can select a higher membership dataset to enhance the SVM classification model. In the experiments, using the proposed new fuzzy selection method which increases the membership can yield better classification rates; and (2) the SVM can effectively determine the structure of patent infringement datasets.

In terms of future work, other types of machine learning datasets with SVMFS would be a challenging issue to study, and one-hot encoding could be considered as a way to input the categorical variables. Future studies could also consider using other data preprocessing techniques to improve the SVM, such as interval PCA [45].

Author Contributions: Conceptualization and methodology: C.-Y.C.; methodology and writing—original draft preparation: K.-P.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science and Technology of the Republic of China, Taiwan, grant number MOST-108-2221-E-029-020.

Conflicts of Interest: The authors declare no conflict of interest.

References

- García-Gil, D.; Luengo, J.; García, S.; Herrera, F. Enabling Smart Data: Noise filtering in Big Data classification. *Inf. Sci.* **2019**, *479*, 135–152.
- van der Gaag, K.J.; Hoogenboom, J.; Busscher, L.; Benschop, C.C.G.; Zuñiga, S.; Sijen, T. The impact of FDSTools noise correction on the analysis of data from the Forenseq™ DNA Signature Prep Kit. *Forensic Sci. Int. Genet. Suppl. Ser.* **2019**, *7*, 797–799.
- Niu, X.; Wang, J. A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting. *Appl. Energy* **2019**, *241*, 519–539.
- Cai, L.; Zhang, Z.; Yang, J.; Yu, Y.; Zhou, T. A noise-immune Kalman filter for short-term traffic flow forecasting. *Physica* **2019**, *536*, 122601.
- Liu, H.; Chen, C. Data processing strategies in wind energy forecasting models and applications: A comprehensive review. *Appl. Energy* **2019**, *249*, 392–408.
- Ma, W.; Qiu, J.; Liu, X.; Xiao, G.; Duan, J.; Chen, B. Unscented Kalman Filter With Generalized Correntropy Loss for Robust Power System Forecasting-Aided State Estimation. *IEEE Trans. Ind. Inform.* **2019**, *15*, 6091–6100.
- Wang, D.; Borthwick, A.G.; He, H.; Wang, Y.; Zhu, J.; Lu, Y.; Xu, P.; Zeng, X.; Wu, J.; Wang, L.; et al. A hybrid wavelet de-noising and Rank-Set Pair Analysis approach for forecasting hydro-meteorological time series. *Environ. Res.* **2018**, *160*, 269–281.
- Florez-Lozano, J.; Caraffini, F.; Parra, C.; Gongora, M. Cooperative and distributed decision-making in a multi-agent perception system for improvised land mines detection. *Inf. Fusion* **2020**, *64*, 32–49.
- Tang, D.; Qin, B.; Wei, F.; Dong, L.; Liu, T.; Zhou, M. A Joint Segmentation and Classification Framework for Sentence Level Sentiment Classification. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1750–1761.
- Jiang, J.-J.; Duan, F.-J.; Wang, X.-Q. An Efficient Classification Method of Mixed Sources. *IEEE Sens. J.* **2016**, *16*, 3731–3734.
- Kasabov, N.K.; Doborjeh, M.G.; Doborjeh, Z.G. Mapping, Learning, Visualization, Classification, and Understanding of fMRI Data in the NeuCube Evolving Spatiotemporal Data Machine of Spiking Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 887–899.
- Shao, J.; Huang, F.; Yang, Q.; Luo, G. Robust Prototype-Based Learning on Data Streams. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 978–991.
- Wang, B.; Zhang, J.; Zhang, Z.; Pan, L.; Xiang, Y.; Xia, D. Noise-Resistant Statistical Traffic Classification. *IEEE Trans. Big Data* **2019**, *5*, 454–466.
- Phan, H.; Andreotti, F.; Cooray, N.; Chén, O.Y.; de Vos, M. Joint Classification and Prediction CNN Framework for Automatic Sleep Stage Classification. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 1285–1296.
- Basha, N.; Sheriffa, M.Z.; Kravaris, C.; Nounou, H.; Nounou, M. Multiclass data classification using fault detection-based techniques. *Comput. Chem. Eng.* **2020**, *136*, 106786.

16. Asuncion, A.; Newman, D. UCI machine learning repository, 2007.
17. Ekong, U.; Lam, H.K.; Xiao, B.; Ouyang, G.; Liu, H.; Chan, K.; Ling, S.H. Classification of epilepsy seizure phase using interval type-2 fuzzy support vector machines. *Neurocomputing* **2016**, *199*, 66–76.
18. Shen, L.; Chen, H.; Yu, Z.; Kang, W.; Zhang, B.; Li, H.; Yang, B.; Liu, D. Evolving support vector machines using fruit fly optimization for medical data classification. *Knowl. Based Syst.* **2016**, *96*, 61–75.
19. Wang, D.; Zhang, X.; Fan, M.; Ye, X. Hierarchical mixing linear support vector machines for nonlinear classification. *Pattern Recognit.* **2016**, *59*, 255–267.
20. Zhong, J.; Tse, P.W.; Wang, D. Novel Bayesian inference on optimal parameters of support vector machines and its application to industrial survey data classification. *Neurocomputing* **2016**, *211*, 159–171.
21. Qi, Z.; Wang, B.; Tian, Y.; Zhang, P. When Ensemble Learning Meets Deep Learning: A New Deep Support Vector Machine for Classification. *Knowl. Based Syst.* **2016**, *107*, 54–60.
22. Liu, Y.; Bi, J.-W.; Fan, Z.-P. A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm. *Inf. Sci.* **2017**, *394*, 38–52.
23. Zhang, X.; Ding, S.; Xue, Y. An improved multiple birth support vector machine for pattern classification. *Neurocomputing* **2017**, *225*, 119–128.
24. Utkin, L.V.; Zhuk, Y.A. An one-class classification support vector machine model by interval-valued training data. *Knowl. Based Syst.* **2017**, *120*, 43–56.
25. Gonzalez-Abril, L.; Angulo, C.; Nuñez, H.; Leal, Y. Handling binary classification problems with a priority class by using Support Vector Machines. *Appl. Soft Comput.* **2017**, *61*, 661–669.
26. Kusakci, A.O.; Ayvaz, B.; Karakaya, E. Towards an autonomous human chromosome classification system using Competitive Support Vector Machines Teams (CSVMT). *Expert Syst. Appl.* **2017**, *86*, 224–234.
27. Richhariy, B.; Tanveer, M. EEG signal classification using universum support vector machine. *Expert Syst. Appl.* **2018**, *106*, 169–182.
28. Ougiaroglou, S.; Diamantaras, K.I.; Evangelidis, G. Exploring the effect of data reduction on Neural Network and Support Vector Machine classification. *Neurocomputing* **2018**, *280*, 101–110.
29. de Lima, M.D.; Costa, N.L.; Barbosa, R. Improvements on least squares twin multi-class classification support vector machine. *Neurocomputing* **2018**, *313*, 196–205.
30. Tang, L.; Tian, Y.; Pardalos, P.M. A novel perspective on multiclass classification: Regular simplex support vector machine. *Inf. Sci.* **2019**, *480*, 324–338.
31. Yang, L.; Dong, H. Robust support vector machine with generalized quantile loss for classification and regression. *Appl. Soft Comput.* **2019**, *81*, 105483.
32. Okwuashi, O.; Ndehedehe, C.E. Deep support vector machine for hyperspectral image classification. *Pattern Recognit.* **2020**, *103*, 107298.
33. Juang, C.-F.; Hsieh, C.-D. Fuzzy c-means bases support vector machine for channel equalisation. *Int. J. Gen. Syst.* **2009**, *38*, 273–289.
34. Yang, X.; Zhang, G.; Lu, J. A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises. *IEEE Trans. Fuzzy Syst.* **2011**, *19*, 105–115.
35. Demidova, L.; Sokolova, Y.; Nikulchev, E. Use of fuzzy clustering algorithms ensemble for SVM classifier development. *Int. Rev. Model. Simul.* **2015**, *8*, 446–457.
36. Eşme, E.; Karlik, B. Fuzzy c-means based support vector machines classifier for perfume recognition. *Appl. Soft Comput.* **2016**, *46*, 452–458.
37. Karlik, B. The positive effects of fuzzy c-means clustering on supervised learning classifiers. *Int. J. Artif. Intell. Expert Syst.* **2016**, *7*, 1–8.
38. Subudhi, S.; Panigrahi, S. Use of fuzzy clustering and support vector machine for detecting fraud in mobile telecommunication networks. *Int. J. Secur. Netw.* **2016**, *11*, 3–11.
39. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Plenum Press: New York, NY, USA, 1981.
40. Dunn, J.C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybern.* **1973**, *3*, 32–57.
41. Jain, A.K.; Dubes, R.C. *Algorithm for Clustering Data*; Prentice-Hall: New Jersey, NJ, USA, 1988.
42. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 1995.
43. Vapnik, V.; Golowich, S.; Smola, A. Support vector machine for function approximation, regression estimation, and signal processing. *Adv. Neural Inf. Process. Syst.* **1996**, *9*, 281–287.

44. Duleba, S.; Farkas, B. Principal Component Analysis of the Potential for Increased Rail Competitiveness in East-Central Europe. *Sustainability* **2019**, *11*, 4181.
45. Aitizem, T.; Bougheloum, W.; Harkat, M.F.; Djeghaba, M. Fault Detection and Isolation Using Interval Principal Component Analysis Methods. *IFAC-Pap.* **2015**, *48*, 1402–1407.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).