Class Specific Feature Selection for Identity Validation using Dynamic Signatures

Dinesh Kumar* and Premith Unikrishnan

Department of Electrical and Computer Engineering, RMIT University, Melbourne, VIC 3000, Australia

Abstract

Classification of the biometrics data for identity validation can be modeled as a single-class problem, where the identity is confirmed by comparing the biometrics of the unknown person with those of the claimed identity. However, current feature selection techniques do not differentiate between single-class and multi-class problems when determining the suitable feature set and select the feature-set that is suitable for representing or discriminating for all the available classes. This may not be the best representation of the biometrics data of an individual because different people may have differences in the most suitable features to represent their biometrical data.

In this paper, a class-specific feature selection method has been proposed and experimentally validated using dynamic signatures. This method is based on the coefficient of variance within the feature set, where the features with smaller variance are selected and the ones with larger variance are rejected. The proposed technique was compared with the other feature selection methods, and the results show that a significant improvement in the classification accuracy, specificity and sensitivity was obtained when using class-specific feature selection.

Keywords: Dynamic signature; Feature selection; Identity validation

Introduction

Accurate validation of the identity requires unique biometric values of the individual. An appropriate feature set to represent the biometrics that will provide small intra-class separation and large inter-class separation is essential. The feature selection for biometrics is achieved by determining the most suitable feature set for the database and using it to define the database. Techniques such as Principal Component Analysis (PCA) [1], and ReliefF [2] are some of the feature selection techniques that are widely accepted. These methods identify the most suitable feature set for representing the database. More recently, features have been identified based on the discriminative ability of the features [3] determined by the outcomes of supervised training of an image, or based on maximum likelihood obtained from the dataset [4]. The fundamental axiom in all these measures is that there is a single template required to represent the entire dataset.

When biometrics is used for validating the identity of an individual, such as for banking purposes, it can be considered as a single class problem, and a single template for representing all the people is not necessary. Each individual can be considered to be a single class, and the different examples of the individual become the database [5]. In such a situation, there can be differences in the features that best represent different classes [6]. This would form the basis for a class specific template. The template of the individual will consist of the unique feature set that best represents the biometrics data of the person, and the associated values. Class Specific Features (CSF) will reduce the intra-class distance and improve the sensitivity of the technique.

In this paper the hypothesis that the dynamic signature biometrics data of different people (classes) may have different set of features that are suitable for compact representation of the class has been tested. While the current biometrics classification method requires a common feature template used to represent the biometrical data for all individuals, this research proposes replacing it by CSF to generate a template with features that are most suitable for representing the biometrics data of the individual. The features would have the least variance between multiple examples of the biometrical data of the individual and would minimize the intra-class distance. We call this technique Class Specific Feature

Selection by Minimizing the Coefficient of Variance (CSFMCOV). The identity of the unknown person would be validated by testing the membership of the biometrical data of the unknown person with the template of the claimed identity. The details of the CSFMCOV and the results of dynamic signature based identity validation using CSFMCOV and other feature selection techniques have been presented in this paper.

Signature based identity validation is widely accepted, and manual verification of signatures by experts has been considered a secure means of identity validation. However the reliability of the use of signature for machine based identity validation has not been well accepted. This can be attributed to two major reasons; (i) large intra-class variability in the shape of the signature of an individual, and (ii) ease with which a person can forge the shape of the signature of the authentic user.

An alternate to signature is the dynamic signature that is based on recording the movement of the stylus rather than the shape of the signature. With easy access to writing tablets and other digital devices that not only capture the shape of the signature, but also the parameters pertaining to the movement such as the velocity, acceleration, and penups, the use of dynamic signatures is now very feasible. Dynamics of the signature are very difficult to imitate because besides the shape of the signature, these have information on the individual strokes and the speed of the different strokes [7]. While the shape of the signature can be copied from an example such as a penned signature, the dynamics are difficult to obtain, and imitate. However, the drawback with the

dynamic signatures is the large number of false negatives, where the signature of the authentic user is not accepted. This is because it is highly unlikely for a person to sign consistently where all the features are repeated and stable [8]. This leads to the difficulty of validating the person based on previous examples, when the repeated examples differ from the training examples.

Dynamic signatures yield very large number of features [7]. Some of the features are the characteristics of an individual, and multiple examples of the signature may be of similar values and the dataset formed would have small variance. However, there are other features of the dynamic signatures that have large variations over repeated experiments and may be considered as distracters. Consider a hypothetical but plausible scenario of one person who may have a very constant speed over repeated examples, but the length of the signature in the horizontal direction may vary over a number of trials, while another person may have highly variable speed but constant horizontal length of the signature. The features that have large variations over repeated signatures could lead to error when being used for confirming the identity of the person and should not be considered to represent the individual. These would be identified as having large variance, while features that have small variance would be suitable for representing the person. However, the currently used feature selection techniques such as principal component analysis (PCA) maximize the variance with the intent of maximizing the inter-class distance. Recent work by Brown et al. [4] has examined the statistical properties of feature selection for properties of datasets. Technique for identifying discriminating feature for segmenting an image that use the supervised training [3] provides a problem specific feature selection method, but has not been used for one-class problem. Biometrics based identity validation is essentially one-class problem where the minimizing the intra-class distance is required, and thus the CSF selection should be based on minimum variance.

Feature Selection

Feature selection for classification of any data is important because presence of some irrelevant features can increase mis-classification and lower the accuracy of classification [9-11]. The other reason is large number of features can make the processing computationally very expensive. Some of the widely accepted feature selection techniques are principal component analysis (PCA) and ReliefF, and number of variations of PCA [1] and ReliefF [2]. An underlying statistical framework has been proposed by Brown et al. [4]. The common factor in all of these techniques is that these identify the most suitable feature set for the entire database consisting of number of classes. However, this is not required for a one-class problem such as biometrics for identity validation. These techniques select the features that have the highest variance [9] and this is useful when the intention is to increase the inter-class distance. However, for single class problem the feature set should have the least intra-class distance, where the suitable features should be repeatable and thus have minimum variation between the multiple examples.

Coefficient of variation (CoV) is a normalized measure of dispersion of a probability distribution and is used as a measure of consistency of the distribution. A small value of CoV indicates that a given distribution has small variations between multiple examples. Small CoV indicates small data scatter and thus the intra-class distance will be small. Based on this, it is proposed that the Class Specific Feature selection by Minimizing the Coefficient of Variance (CSFMCOV) would lead to more accurate biometrics classification for identity validation.

In a probability density based classifier like Bayesian classifier it can be observed that as the classification performance is affected by large variances in data.

In Naïve Bayesian Classification, the probability of a given unknown feature f to belong to a class c is given by equation (1)

$$P(f \mid c) = \frac{1}{\sqrt{2\Pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}} \tag{1}$$

Where μ is the mean of the feature and σ is the standard deviation of the feature in the training set.

To determine the efficacy of the proposed technique, the features identified using CSFMCOV were compared with those obtained by commonly used and widely accepted feature selection techniques (i) Principal component analysis (PCA) [1], ReliefF [2], Gain Ratio [12,13], Information gain [12,13], Gini Index [13], and Sequential Forward search [14].

Classification

One-class support vector machine (OSVM)

Biometrics based identity validation requires the testing of the membership of the unknown person against the claimed class, and this is a one-class problem [15]. Unlike many other classification problems, the training data for identity validation may only contain the examples of the authentic user (positive class), and not have the data pertaining to the negative class. Even when the negative data may be available, it would not be representative of the entire negative class. Traditional classifiers are not suitable for such an application because these require examples from both classes, and generally assume that the dataset is balanced.

One-class Support Vector Machine (OSVM) was first proposed by [16]. It is an extension of the original two class SVM. OSVM enables the training of the classifier in the absence of any negative example data. During the training phase, a certain user-defined fraction of the positive class data-points are treated as if they belong to the negative class.

The LIBSVM implementation [9] was selected for this research.

Given training vectors $x_i \in R^n, i=1,..l$ without any class information, the primal problem of One-Class SVM is

$$Min_{w,\varepsilon,p} \frac{1}{2}\omega^T\omega - \rho + \frac{1}{vl}\sum_{i=1}^{l}\varepsilon_i$$

subject to $\omega^T\varphi(x_i) \geq \rho - \in_i, \in_i \geq 0, i=1,...,l$

The dual problem is

$$Min_\alpha = \frac{1}{2}\alpha^T Q\alpha \tag{2}$$

Subject to $0 \leq \alpha_i \leq \frac{1}{vl}, i=1,....,l$

$e^T\alpha = 1$

Above $v \in [0,1]$ is an upper bound on the fraction of margin errors and a lower bound on the fraction of support vectors, (ω, ρ) are a weight vector and an offset parameterizing a hyper plane in the feature space associated with the kernel, $l \in N$ is the number of observations. φ is a feature map $X \to F$, a map into an inner product space F such that the

inner product in the image of φ can be computed by evaluating some simple kernel and α is a multiplier $\alpha \geq 0$. ε_i is called a slack variable which measures the degree of misclassification of x_i. $e = [1,....,l]^T$ is the vector of all positive classes,

Where $Q_{ij} = K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ The decision function is

$$\text{sgn}\left(\sum_{i=1}^{l} \alpha_i K(x_i, x) - \rho\right)$$

Solving (2) by OSVM implementation in LIBSVM is equivalent to solving

$$Min_\alpha = \frac{1}{2}\alpha^T Q\alpha \tag{3}$$

Subject to $0 \leq \alpha_i \leq \frac{1}{vl}, i = 1,....,l$

$$e^T \alpha = vl$$

The trained OSVM and the feature template for each individual were saved for validating the identity of an individual. To validate the identity, the values of the biometrical features of the unknown person corresponding to the feature set template of the claimed identity were the input to the OSVM of the claimed identity.

Experimental Setup

A portable Graphic Tablet with no visual feedback was used to acquire the signatures from the users. The x and y coordinates, the time, t, and the pressure, p, corresponding to the touch point of the stylus on the tablet were captured at the rate of 100 samples/second. The table with four columns t, x, y and p corresponding to each signature example was saved as a text file. The recordings were manually segmented by the examiner to identify the start and completion of each recording.

Fifty-five adults volunteered to participate in this study. The exclusion criteria were (i) pain in the dominant hand, (ii) tremor in the upper extremity and (iii) consumption of any intoxicant within the past 24 hours. During the experiment, the participants were either seated in an office chair in their own surroundings/environment or in the laboratory. The participants were given demonstrations and some time to help them become accustomed to the tablet prior to the actual recordings.

Ten signatures were recorded from each participant. To ensure that the recordings represented the diversity of the individual, two sets of five signatures were recorded, with a minimum of one-hour delay between the two recordings. A total of 550 signatures were collected from the 55 participants. Preliminary experiments were also conducted using a smart-phone instead of the digital tablet, which gave a visual feedback to the user. While these results have not been reported in this paper, these were very similar to the results obtained using the tablet. However, users found the smaller size to be less convenient, and thus was not used.

Data Analysis

Feature extraction

Each recording was segmented manually to identify the start and the end of each signature. Based on the work reported by [17-19], a set of thirty features were obtained. These have been listed below (Table 1).

Feature selection

The minimum number of components (features) n to represent the dataset was obtained using PCA, and corresponding to 99% cumulative

Penup	Theta (Mean path tangent Angle)	Duration
Speed	Vn (Mean path velocity magnitude)	Mean-Pressure
MinVx/AvgVx (V_x=horizontal velocity)	Pn (Mean Log curvature radius)	Duration Vx>0/Duration
MinVy/AvgVy (V_y=vertical velocity)	An (Mean Total acceleration magnitude)	Duration Vy>0/Duration
V/Vmax	Pen-DownRatioTime	Total Vx>0/Duration
SDVelocity (standard deviation of velocity)	Cursiveness	Total Vy>0/Duration
AccelerationX	TopHeaviness	Total Vx=0
AccelerationY	BottomHeaviness	Total Vy=0
Direction Change-x	Ratio l/h (total length/total height)	Vavg/MaxVx
Directionchange-y	Ratio Area/Distance	Vavg/MaxVy

Table 1: List of the features of dynamic signature.

energy. Based on the PCA results, five features (n=5) were considered to be suitable for representing the dataset. The widely available and accepted feature identification techniques; (1) PCA, (2) Gini, (3) Information Gain (IG), (4) ReliefF, (5) Sequential Forward Search (SFS) and (6) Gain ratio were used to identify a minimum number of features to give the best classification result. In Gini, IG, GainRatio and Relieff feature selection methods, five features were found to be sufficient to give a minimum classification error after which the error stabilized with increase in number of features, and the error increased when the number of features were more than twenty-two. In SFS the error rate stabilized from feature number four to fourteen before increasing again. To maintain uniformity we selected best five features to compare between different methods. This was repeated for CSFMCOV and repeated for each class (individual). Orange data mining software [20] was used to obtain the features for methods ReliefF, Gain Ratio, Gini and IG while software was developed on Matlab platform for the remaining feature selection techniques.

Classifier training and validation

Monte Carlo Analysis: To test the system and determine the specificity, sensitivity and repeatability, Monte-Carlo simulation method was used to generate 100 samples using the distribution of the recorded data and the 1st and 2nd order statistics (mean, μ, and standard deviation, σ). This simulated data was made the input to the trained classifiers and the mean classification error was obtained.

Cross validation: The OSVM was trained for each class (person) using the five features as the input and the default value of the parameter v=0.1 was chosen. Ten-cross, leave-one-out approach [21] was used to validate the system which is suitable when the number of examples is small. The OSVM was trained using nine examples, and tested using the tenth example corresponding to the positive class and with the remaining 540 samples obtained from other participants corresponding to the negative class. Each class was tested ten times such that each example from the set of ten was used to test the system. The total number of trials using this strategy was 550 with 550 possible true positives and 297,000 (540*550) true negatives which can also be considered as cases of random forgery i.e. one true positive and 540 true negatives for each trial. Accuracy, sensitivity and specificity were computed for each class, and the average of all the classes was then calculated. An rbf kernel was used. The kernel density function for an RBF kernel is:

$$K(X_i, X_j) = \exp(-g*(X_i - X_j)^2, g > 0$$

To obtain the Receiver Operating Characteristics (ROC) plots, the training was repeated for four values of kernel hyper parameter gamma, g ;0.0001, 0.001, 0.01 and 0.1.

Results

Table 2 is the set of highest ranked five features identified for the dynamic signature database using five widely accepted feature selection methods. The list of features selected using the proposed CSFMCOV cannot be shown in table 1 because this technique identifies different features for each person. The results show that while there is similarity between the features selected using Gini, Gain Ratio and Information Gain, there is a large discrepancy between the features selected using other feature selection methods. This indicates that the selection of features is dependent on the method.

Figure 1 shows the frequency of occurrence of the class specific features using CSFMCOV. The *x*-axis shows the features numbered one till thirty and the *y*-axis shows the frequency of selection of each feature for the different classes. From this figure, it is observed that there is a wide spread of the feature selection, and wide variation in the frequency of occurrence of different features. This confirms the hypothesis that the dynamic signatures of different people have different set of features that are most suitable for representing their examples.

The accuracy, sensitivity and specificity of cross-validation using OSVM have been tabulated in table 3. There are seven rows, with each row corresponding to the different feature selection methods. From this table, it is observed that CSFMCOV has the highest accuracy, specificity, and sensitivity compared with all the other techniques.

Figure 2 is the plot of the number of features and the mean error calculated using the leave one out method for OSVM after Monte Carlo simulation for all classes in correctly identifying the person. The results show that error reduced as the number of features were increased from

Feature selection method	List of five highest ranked features				
Gini	Penups	Duration	Cursiveness	Speed	SdVelocity
InfoGain	Penups	Duration	Cursiveness	Speed	SdVelocity
ReliefF	Penups	Pen-Down RatioTime	VelocityX	Duration	Speed
SFS	Penups	Duration	Direction Change-x	Cursiveness	Ratio Area/ Distance
Gain Ratio	Penups	Duration	Cursiveness	Speed	SdVelocity

Table 2: List of highest five ranked features using five widely used feature selection techniques.
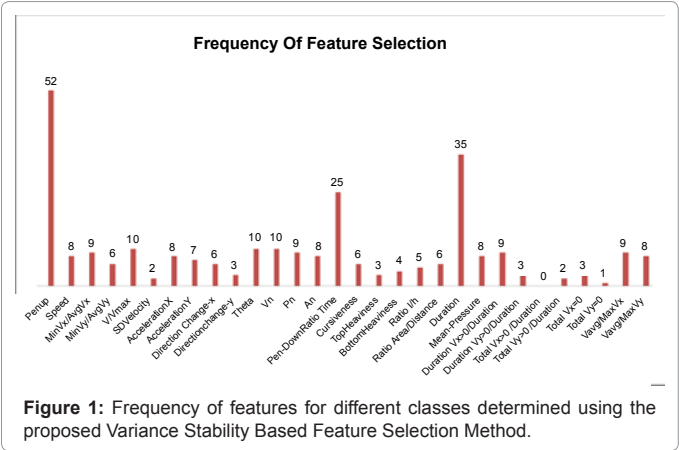

Figure 1: Frequency of features for different classes determined using the proposed Variance Stability Based Feature Selection Method.

Feature selection method	Accuracy	Sensitivity	Specificity
Gini	0.998656	0.867272	0.999969
Info Gain	0.998656	0.867272	0.999969
PCA	0.999815	0.918181	0.99996
Relieff	0.999829	0.927272	0.999962
SFS	0.999734	0.869090	0.999976
CSFMCOV	0.999976	0.996363	0.999983
GainRatio	0.998656	0.867272	0.999969

Table 3: Average classification accuracy, sensitivity and specificity based on leave-one-out using features using the seven feature selection techniques.
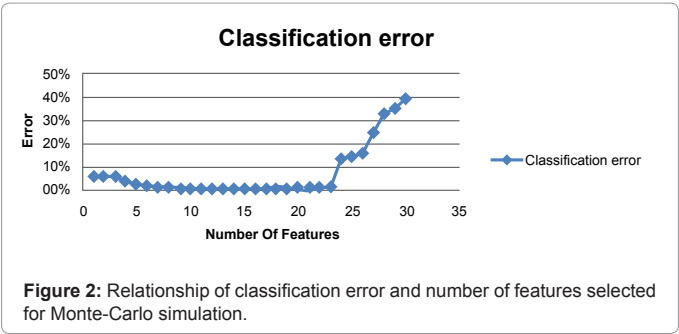

Figure 2: Relationship of classification error and number of features selected for Monte-Carlo simulation.
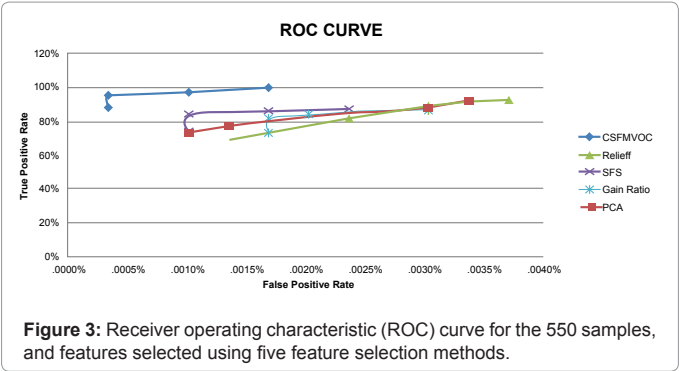

Figure 3: Receiver operating characteristic (ROC) curve for the 550 samples, and features selected using five feature selection methods.

1 till 5, remained stable from 6 till 20, and then increased again. This indicates that a minimum of 5 features is enough for good accuracy when using dynamic features.

Figure 3 is the ROC plot based on the four values of threshold, *g*. Each curve corresponds to the different feature selection method. From this figure, it is observed that ROC corresponding to CSFMCOV covers a higher area.

Discussions

Dynamic signature as the biometrics for identity validation is a one-class problem and this does not require a common template for all the individuals. In such a situation, each class can be represented by a feature set that best represents it, and these features would be the basis for its template. This forms the basis for class-specific feature selection.

The results of this work have demonstrated there is a big diversity in the features for selected using different feature selection methods for representing the biometrics data, showing that the feature selection is dependent on the choice of the method. The proposed CSFMCOV selects only those features that have small variance among multiple examples for a specific class. From figure 1, it is observed that there is a broad range of features selected for the dynamic signatures of different

people by CSFMCOV. This indicates that there is no unique feature set for dynamic signature of different people.

The results clearly indicates that while specificity of all the feature selection methods are good, the sensitivity is significantly better when class-specific features were selected using CSFMCOV. This is particularly relevant features of dynamic signatures as they are susceptible to huge intra-class variances leading to poor sensitivity. This study shows that the number of positive classes rejected by a classification model is greatly reduced by choosing a class-specific feature selection model. There is no other reported class-specific feature selection technique and thus CSFMCOV cannot be compared with other similar technique. However, recently reported discriminative feature selection [21] has in-principle similarities with CSFMCOV, even though it has been used very differently. The results show that dynamic signature, when used with CSFMCOV, is accurate, sensitive and specific for identity validation. CSFMCOV has overcome one of the earlier shortcomings of low sensitivity, and with the easy access to digital tablets, number of applications of this identity validation technique are possible. However, it is important to note that CSFMCOV would not be suitable if the problem was of identification rather than validation because in such a situation, a common template for all classes would be necessary.

The results show that the suitable features for a one-class classification problem can be selected based on the variance, with the features with smaller variance being more suitable than features with larger variance. This is in contrast to feature selection methods such as PCA that identify the most suitable features based on maximizing the variance [22]. This difference is because other feature identification methods maximize the inter-class distance while CSFMCOV is based on minimizing the intra-class distance. This is suitable for applications where there is inter class variances between features along with high variances in intra class features.

This research has identified the features based on the variance of the sample data. Though these results were excellent, it is possible that presence of outliers may lead to sub-optimum results because these can distort the value of mean and standard-deviation. Other statistical measures such as the use of range or median measure or neural network approach may also give similar results, and overcome issues such as due to outliers.

Conclusion

Biometrics for identity validation is one-class classification problem, and does not require a global feature set. The accuracy, specificity and sensitivity are better if the feature set is specific to the class, and this is achieved by identifying the features that have small variance. This is different from multi-class problems where it is essential to use a global feature set for all the classes, and often the selected features have highest variance. One-class classification problems requires narrowness of the data points belonging to the class which leads to small intra-class distance, while the multi-class problem requires the maximizing of the inter-class distance.

This paper reports the use of minimizing the coefficient of variation for feature selection, and is called Class Specific Feature selection by Minimizing the Variance (CSFMCOV). However, the authors acknowledge that there may be other methods to identify the Class Specific Features (CSF), such as the use of neural networks or other similar algorithms. The major outcome of this work is determining the role of CSF for biometrics and other one-class problems. While this paper has tested the application of CSFMCOV for dynamic signatures for identity validation, this may be suitable for many other one-class problems.

References

1. Krawczyk S (2005) User authentication using on-line signature and speech. Master's Thesis, Department of Computer Science and Engineering, Michigan State University.

2. Robnik-Sikonja M, Kononenko I (2003) Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learning Journal 53: 23-69.

3. Shang L, Xiao B (2012) Discriminative features for image classification and retrieval. Pattern Recognit Lett 33: 744-751.

4. Brown G, Pocock A, Zhao MJ, Lujan M (2012) Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. J Mach Learn Res 13: 27-66.

5. Jain AK, Griess FD, Connell SD (2002) On-line signature verification. Pattern Recognit 35: 2963-2972.

6. de Lannoy, François D, Verleysen M (2011) Class-Specific Feature Selection for One-Against-All Multiclass SVMs. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges.

7. Plamondon R, Srihari SN (2000) On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. IEEE Trans Pattern Anal Mach Intell 22: 63-84.

8. Guest RM (2004) The repeatability of Signatures. IWFHR '04 Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition, Tokyo.

9. Dash M, Liu H (1997) Feature selection for classification. Intelligent Data Analysis 1: 131-156.

10. Janecek AGK (2008) On the Relationship Between Feature Selection and Classification Accuracy. J Mach Learn Res 4: 90-105.

11. Navot A (2006) On the Role of Feature Selection in Machine Learning. PhD Thesis, Hebrew University.

12. Grimaldi M, Cunningham P, Kokaram A (2003) An evaluation of alternative feature selection strategies and ensemble techniques for classifying music. The 14th European Conference on Machine Learning and the Seventh European Conference on Principles and Practice of Knowledge Discovery in Databases, Dubrovnik, Croatia.

13. Li T, Zhang C, Oghiara M (2004) A comparative study on feature selection and multiclass classification methods for tissue classification based on gene expression. Bioinformatics 20: 2429-2437.

14. Pudil P, Novovicova J, Kittler J (1994) Floating search methods in feature selection. Pattern Recognit Lett 15: 1119-1125.

15. Tax DMJ (2001) One-class classification-concept learning in the absence of counterexamples. PhD Dissertation, Delft University of Technology, ASCI Dissertation Series 65: 1-190.

16. Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC (1999) Estimating the support for a high-dimensional distribution. Microsoft Research, WA, USA.

17. Lee LL, Berger T, Aviczer E (1996) Reliable on-line human signature verification systems. IEEE Trans Pattern Anal Mach Intell 18: 643–647.

18. McCabe A, Trevathan J, Read W (2008) Neural Network Based Handwritten Signature Verification. Journal of Computers 3: 9-22.

19. Nelson W, Kishon E (1991) Use of dynamic features for signature verification. IEEE International Conference on Systems, Man and Cybernetics.

20. Demšar J, Zupan B (2004) Orange: From experimental machine learning to interactive data mining.

21. Joachims T (2000) Estimating the generalization performance of a SVM efficiently. ICML'00 Proceedings of the Seventeenth International Conference on Machine Learning, San Francisco 431-438.

22. Chang CC, Lin CJ (2001) LIBSVM: A Library for Support Vector Machines.