Paper

On High Generalization Ability of Test Feature Classifiers

Non-member Vakhtang Lashkia (Okayama University of Science)
Member Shun'ichi Kaneko (Hokkaido University)
Non-member Mitsuru Okura (Okayama University of Science)

Test feature classifiers are generated directly from training samples and have a 100% recognition rate on training data. Although this perfect learnability is an important feature of the classifiers, it does not guarantee a good generalization. In this paper, we concentrate on the performance of classifiers on test data, and describe cases when a 100% recognition rate can be achieved. We show that training data can contain information about possible discriminant boundaries between entire classes. In general, it is impossible to extract this information, although we propose a heuristic algorithm which could lead to a 100% recognition rate. To test the performance of the classifiers, we apply them to both artificial and real data. For the real data, we use the well-known breast cancer and satellite image databases. Our experimental results show that the proposed classifiers have not only a high recognition ability, but also confirm the ability of a 100% recognition rate in real classification problems.

Keywords: nonparametric classifier, learning, classification, test feature, breast cancer database, satimage database.

1. Introduction

Classification techniques are one of the most important subjects in the field of pattern analysis. The goal of designing pattern classification systems is to achieve the best possible classification performance for the task at hand. There are several factors to be considered. In many applications, the most important factor becomes the ability of the classifier to exhibit a good generalization.

In this paper, we present test feature classifiers and concentrate on their generalization ability. These classifiers are generated directly from training samples using so-called *tests*, sets of features that are sufficient to distinguish patterns from different classes of training samples (do not confuse with test data). The concept of the test was first introduced in Chegis & Yablonsky [1] for the purpose of digital logic circuit analysis. The first use of tests, as the pattern recognition tool was reported in Zhuravlov et al. [2], and several algorithms for test extraction, which are asymptotically best (i.e. when the number of features is large, the algorithm selects only tests in each step and does not make idle steps) were proposed in the papers [3] [4]. Many theoretical aspects of test feature classifiers including an estimation of the number of all tests were considered in Aleshin [5]. The real applications of test feature classifiers are presented [14]~[16]. In Lashkia et al. [14], an application of test feature classifiers to textual region location was considered. In Itqon et al. [15], some extensions of test feature classifiers were introduced and an application to character recognition was considered. Test feature classifiers were also applied to the phoneme database [16]. In the above applications a high generalization ability was achieved. These experimental results were very encouraging and showed the importance of further investigation of test feature classifiers.

In this paper we address some issues relevant to the ability of test feature classifiers to have a high recognition rate. In some experiments [14] [16] there were cases when test feature classifiers achieved a 100% recognition rate. Our purpose in this paper is to understand why and when a 100% recognition rate is possible and propose an algorithm which can lead to a perfect generalization.

By extracting tests from a training set, test feature classifiers guarantee to have a 100% recognition rate on training data. Although many classifiers can also learn perfectly training data, this does not guarantee a good recognition rate on test data. We show that in the case of test feature classifiers the set of tests can also contain information about possible discrimination boundaries between entire classes, which we aim to recognize. If we are able to extract this information by extracting *kernels* from tests we will obtain a 100% recognition rate on any test data. In general, it is impossible to find this set of kernels but we propose a heuristic kernel detection algorithm, employ it in real applications, and show its effectiveness.

We apply the proposed classifiers to both simulated and real data. In the real applications, we use the well-known breast cancer and satellite image databases. There are many papers related to experiments on these databases [6]~[12]. Among them, comprehensive studies of instance-based learning algorithms are presented [6] [7]. Most of instance-based learning algorithms are based on metrics. Although these algorithms perform well and are one of the most reliable classifiers in real-world domains, they have difficulty to achieving high recognition rates when classes become overlapped. The misclassifications of the near-boundary instances often reduce performance. Experiments [6]~[9], [11], [12] show

that distance based classifiers have the best performance on the breast cancer and satellite image databases, although there performance results are still unsatisfactory. Classifiers based on optimization methods are also very popular and often appear in real-world applications. They too, have difficulties in achieving high recognition rates when there are complicated overlapped classes. The optimization procedures terminate when a set of parameters that correctly classify the training data is found. However there could be many decision boundaries which hold global extrema and correctly separate training data. The selected decision boundary also may not correspond to an intuitive notion of what constitutes a good decision boundary [13]. In the case of the breast cancer and satellite image databases, classifiers based on optimization methods perform worse than distance based classifiers [9], [12]. Other types of classifiers such as statistical parametric classifiers [11], [12] and decision trees [9], [12] show even worse recognition results.

Our experiments on real data confirm theoretical ability of test feature classifiers to have a 100% recognition rate. The proposed kernel detection algorithm give good result in kernel detection. On the breast cancer database a 100% recognition rate was achieved with a relatively small number of rejections. On the satellite image database, a 100% recognition rate was obtained with a very large number of rejections. In Section 3 we indicate possible was to reduce the number of rejections and show that the ideal way to achieve complete classification without rejections is by constructing a *prototype* training set. Despite the number of rejections, a 100% recognition rate is very important, especially in applications where mistakes could cause fatal results.

This paper is organized as follows: in Section 2, we introduce test feature classifiers and in Section 3 we discuss their properties and performance. In Section 4, we present results on simulated and real data.

2. Basic Concept and Notations

Assume that P is an n-dimensional feature space, $P = \{\boldsymbol{t} = (t_1, ..., t_n)\}$, and each pattern is represented as a binary-valued feature vector in this space $t_i \in \{0, 1\}$. Let us also assume that there are two possible classes I_1 and I_2. The problem of designing a classifier for pattern recognition can be stated as follows: a function V must be found such that a pattern $\boldsymbol{x}$ is in the class I_1 (in the class I_2) if and only if $V(\boldsymbol{x}) \geq 0$ ($V(\boldsymbol{x}) < 0$).

Let us denote $B_1 = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^{m_1}\}$ as a set of training samples from the class I_1 and $B_2 = \{\boldsymbol{y}^1, ..., \boldsymbol{y}^{m_2}\}$ as a set of training samples from the class I_2, where $\boldsymbol{x}^j = (x_1^j, ..., x_n^j), j = 1, ..., m_1$, $\boldsymbol{y}^j = (y_1^j, ..., y_n^j), j = 1, ..., m_2$, and $I_1 \cap I_2 = \emptyset$. A collection of k features $(1 \leq k \leq n)$,

$$\tau = \{i_1, ..., i_k\}$$

is called a *test feature* (or test) of B_1 and B_2 if for any p $(1 \leq p \leq m_1)$ and any q $(1 \leq q \leq m_2)$ there exist some $i_s \in \tau$ $(1 \leq s \leq k)$ such that $x_{i_s}^p \neq y_{i_s}^q$. In other words, a test is a collection of features which is sufficient to distinguish vectors from different classes of training samples. If for a test τ, the set $\tau - \{i_s\}$ is not a test for any s $(1 \leq s \leq k)$, then τ is called a *prime test feature* (or prime test).

It is important to note that the assumption of binary-valued features in the test definition is not essential. A test could be defined similarly on a many-valued or real-valued feature space. The only requirement is to have an inequality relation R on the feature space, such as an ordinary inequality $\neq$, or any other specific inequality. The definition of a test is formulated as follows. A collection of features, $\tau = \{i_1, ..., i_k\}, (1 \leq k \leq n)$ is called a test of B_1 and B_2 if for any p $(1 \leq p \leq m_1)$ and any q $(1 \leq q \leq m_2)$ there exist some $i_s \in \tau$ $(1 \leq s \leq k)$ such that $x_{i_s}^p \, R \, y_{i_s}^q$. Many-valued and real-valued cases can be easily reduced to the binary cases, therefore we concentrate on the basic binary case below.

A test $\tau = \{i_1, ..., i_k\}$ can be considered as an n-tuple vector,

$$\boldsymbol{\tau} = (\tau_1, ..., \tau_n)$$

where τ_i is 1 if $i \in \{i_1, ..., i_k\}$, and 0 otherwise. Denote the number of features (1s) in a test $\boldsymbol{\tau}$ as $|\boldsymbol{\tau}|$. We say $|\boldsymbol{\tau}|$ is the length of $\boldsymbol{\tau}$. A test is a collection of features for discriminating training samples of different classes and it can be used for the classification of unknown patterns.

For a given test $\boldsymbol{\tau}$ we can measure the degree of similarity of an unknown pattern $\boldsymbol{t}$ to the training pattern $\boldsymbol{x}$ by

$$\prod_{i=1}^{n}(1 - \tau_i|t_i - x_i|) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (1)$$

This expression takes the value 1 if and only if $\boldsymbol{t}$ and $\boldsymbol{x}$ coincide in the features defined by test $\boldsymbol{\tau}$, and takes the value 0 otherwise. In this case, no metric is used. The coincidence of $\boldsymbol{t}$ and $\boldsymbol{x}$ defined by test $\boldsymbol{\tau}$ can be expressed as $\boldsymbol{\tau} \circ \boldsymbol{t} = \boldsymbol{\tau} \circ \boldsymbol{x}$, where the symbol $\circ$ means a projection operator such that $(a_1, a_2, \cdots, a_n) \circ (b_1, b_2, \cdots, b_n) = (a_1 \cdot b_1, a_2 \cdot b_2, \cdots, a_n \cdot b_n)$.

Let T be a set of tests. Taking (1) as a measure of similarity we calculate votes $V_1(\boldsymbol{t})$ and $V_2(\boldsymbol{t})$ for the classes I_1 and I_2 in the following way

$$V_1(\boldsymbol{t}) = \frac{1}{m_1} \sum_{\boldsymbol{\tau} \in T} \sum_{\boldsymbol{x} \in B_1} \prod_{i=1}^{n}(1 - \tau_i|t_i - x_i|)$$

$$V_2(\boldsymbol{t}) = \frac{1}{m_2} \sum_{\boldsymbol{\tau} \in T} \sum_{\boldsymbol{y} \in B_2} \prod_{i=1}^{n}(1 - \tau_i|t_i - y_i|).$$

We call a classifier based on the discriminant function

$$V(\boldsymbol{t}) = V_1(\boldsymbol{t}) - V_2(\boldsymbol{t})$$

as *test feature classifier* with T and denote it by TF_T [5]. We extend TF to reject patterns $\boldsymbol{t}$ for which $V(\boldsymbol{t}) = 0$, and denote TF_T classifier as TFR_T for the optional function of rejection.

3. Properties and Performance

Since V is a polynomial of n variables with a degree

less than or equal to n [16], for large values of n, the evolution of TF becomes time consuming. We can improve the time requirement by extracting important (for classification) features and reducing the dimension of a feature space. Denote $\hat{T}$ as the set of all prime tests, and $\hat{T}_i, i = 1, ..., n$, as the set of all prime tests containing the ith feature. We define a info vector as

$$\boldsymbol{p} = (p_1, ..., p_n)$$

where $p_i = |\hat{T}_i|/|\hat{T}|$ is an info weight [5]. We assume that the more the prime tests contain the ith feature, the more the ith feature is important for the classification purpose. The info weight can be considered as a measure of the feature importance. Therefore, features can be sorted by their importance and we can reduce their number by removing features with small info weights. Because the construction of the set of all prime tests is very time consuming (there are sets with even an exponential number of prime tests), in our experiments to calculate the info vector we use a set of short prime tests instead of $\hat{T}$.

It is easy to prove that for any training set B_1 and B_2 ($B_1 \cap B_2 = \emptyset$), the classifiers TF and TFR have no error on the training samples. As seen from the definition of the test feature classifier, the classification performance on the test samples depends on the set of tests T, and on the set of training samples B_1 and B_2. Let us call a test (a prime test) of I_1 and I_2 a *kernel* (a prime kernel). Suppose that there exists a non-trivial (different from $(1, 1, ..., 1)$) set of kernels. If $\boldsymbol{\kappa}$ is a kernel then obviously $\boldsymbol{\kappa}$ is a test for B_1 and B_2 and the following relation holds

$$\boldsymbol{\kappa} \in \bigcap_{B_1 \in I_1, B_2 \in I_2} T \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (2)$$

where T is the set of all tests for B_1 and B_2. Suppose that a set of kernels K for unknown I_1 and I_2 is found. It is easy to see that TFR_K has a 100% recognition rate on any test samples for any training set. Each set of kernels defines a discriminant boundary between the classes I_1 and I_2. By extracting tests from the training set, we are guaranteed to have a 100% recognition rate on the training data. Then, by extracting kernels from the set of tests, we are guaranteed to have a 100% recognition rate on test data. In general, it is impossible to find the set of kernels for unknown I_1 and I_2, but we can estimate it from the training sets using relation (2), or we can construct some heuristical algorithms for kernel detection.

Suppose that l is the minimal length of tests. Let us propose the following simple heuristic kernel detection algorithm.

First, the set T containing all tests with a length no more than d is formed. Next, an info vector is calculated using T. Since the length of the kernel is more or equal to the length of the minimal test, the length k of the candidate kernel is selected as $k \geq l$. Finally, by choosing k features with highest info weights we construct a candidate kernel.

The value d is a parameter of the algorithm and it is determined experimentally depending on the available computational power necessary for tests detection. If d is small, it is easier to detect tests, but a more reliable estimation of the info vector can be obtained if d is large. A preferable compromise for d is the value $l + 1$, which we use in our experiments on real data.

If we detect a set of kernels K, TFR_K gives a 100% recognition rate on any test samples. But the TFR classifier can also reject samples when the discriminant function is 0, and this can happen in many cases. For TF classifier even if we find a set of kernels we need an appropriate training set to obtain a recognition rate of 100%. We say that the pair (B, K), $B = B_1 \cup B_2$, covers a pattern $\boldsymbol{z}$ if there exist $\boldsymbol{x} \in B$ and $\boldsymbol{\kappa} \in K$ such that $\boldsymbol{z} \circ \boldsymbol{\kappa} = \boldsymbol{x} \circ \boldsymbol{\kappa}$. Denote by $C(B, K)$ the set of all $\boldsymbol{z}$ that are covered by (B, K). We call a set B a *prototype* set for K if $C(B, K) \supseteq I_1 \cup I_2$. It is easy to see that if B is a prototype set for K then TF_K will have a 100% recognition rate on any test samples.

The proposed kernel detection algorithm detects only one candidate kernel for each training set. In the cases when a prototype set is not available, test feature classifiers based on only one kernel will lead to a large number of rejections. However, despite this number of rejections, the success in kernel detection gives a 100% recognition rate. This is very important, especially in applications where mistakes could cause fatal results. We apply this algorithm to the real data in Section 3, and show its effectiveness.

It can be proved that I_1 and I_2 have only one prime kernel when $I_1 \cup I_2 = P$. If the set T consists only of the test $(1, 1, ..., 1)$ then test feature classifier is degenerated and becomes useless. Test feature classifier works well for the cases when $I_1 \cup I_2$ is a small part of P and has many kernels. These are cases which we encounter in reality when n becomes large.

4. Experiments

4.1 Artificial Data Several experiments on artificial data were conducted in the previous works [14], [15]. In this paper let us consider and analyze the experiment on one of the artificial data [14].

We consider the problem with two adjacent classes of two-dimensional region in a 32×32 square shown in Fig. 1. The class 1 is represented by the gray area and the class 2 by the white area. The coordinates of patterns (points) are utilized as features. The number of randomly chosen training patterns was varied from 30 to 100, and 1024 patterns are used for testing.

The performance of TF was compared with one of the most popular non-parametric classifiers, a single nearest neighbor classifier (NN). Both NN and TF classifiers have no error on the training samples. A NN classifier with Euclidean distance (NN_1) is applied directly to two-dimensional patterns. For TF, a 10 bit binary representation of coordinates (each coordinate is represented by five bits) was used and the set of all prime tests of the training samples was employed. A NN classifier with Hamming distance (NN_2) is applied to the

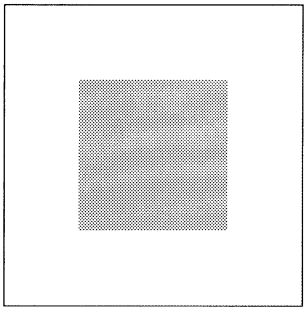

Fig. 1. A two-class problem with a kernel (1100011000).

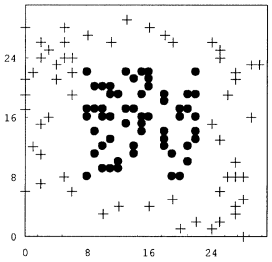

Fig. 2. Examples of training sample. 60 black circles show the class 1 samples and 60 crosses the class 2.

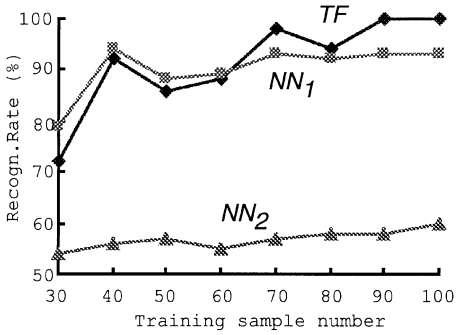

Fig. 3. Recognition rate vs training sample number

Table 1. Template patterns in each class

class	feature patterns
1	01--- 01---
	01--- 10---
	10--- 01---
	10--- 10---
2	00--- -----
	01--- 00---
	01--- 11---
	10--- 00---
	10--- 11---
	11--- -----

'-' means 'don't care'.

same binary data as *TF*.

The two-class problem shown in Fig. 1 has the prime kernel (1100011000). Since the classes can be constructed from combination of quarters in both axes (features), upper two bits in each feature are enough to distinguish the classes as shown in Table 1. For this kernel, 16 prototypes are needed to have a 100% recognition rate on any test samples. When the training set contained 80, 90, or 100 samples, the set of all detected prime tests coincided with the kernel. However only the cases with 90 and 100 training samples gave a 100% recognition rate as shown Fig. 3. This is because in these cases a prototype set was contained in the training set. This and other experiments show that kernels can be detected from training sets and confirm the ability of a 100% recognition rate on a test data.

In the presented experiment, $I_1 \cup I_2 = 2^n$, where $n = 10$. As we mentioned at the end of Section 3, in such cases there is only one kernel. In real applications,we usually encounter a large number of kernels [14], which makes test feature classifiers very useful and practical.

4.2 Real Data A number of experiments are conducted on real data. For the real data, we use the well-known breast cancer and satellite image (satimage) databases, which are available via ftp at ftp.dice.ucl.ac.be. In order to have a reasonable estimation of the performance of the classifiers we use the same holdout method with two trials as in Woods et al. [12]. Each data set is randomly partitioned into two equal halves, keeping the class distributions similar to that of full data set. Initially, one set is used as training data, and the classification accuracy is evaluated using the other set. Next, the roles of the two sets are reversed.

The classification problem of the breast cancer data is to distinguish between benign and malignant. This database is composed of two classes in nine dimensions. Each feature is integer-valued and varies from 1 to 10. There are 699 samples in the database. In Wilson & Martinez [6], where 90% of the data samples were used as a training and remaining 10% as testing data, the best recognition result show by the Edited Nearest Neighbor, a 97% recognition rate. In Wolberg et al. [9] and Mangasarian et al. [10], experiments were conducted on 369 samples. 50% of samples were used as a training data, and on the remaining 50% of samples a 93.5% recognition rate was achieved.

We removed 16 samples with missing attribute values from the breast cancer database and test feature classifiers are evaluated on the rest of the 683 samples. We apply test feature classifiers directly to the nine dimensional integer-valued breast cancer database. The inequality R is defined as, $xR\,y$ if and only if $|x-y| > tr$, where $tr = 1$. The value 1 is the maximal value of tr which still makes sense of the use of test approach.

In each trial, the kernel detection algorithm is applied to the training samples. First, the set T containing all tests with length no more than $d = l + 1$ (where $l = 3$ is a minimal length of the detected tests) is constructed. Then, by calculating an info vector from the set T, we construct a candidate kernel τ of length d , by taking the d features with highest info weights. In the first trial, $|T| = 28$. Fig. 4 shows distributions of the info vector of a training set. The proposed kernel is $\boldsymbol{\tau} = (101011000)$, which corresponds to 4 higher values of the info weights. In the second trial,

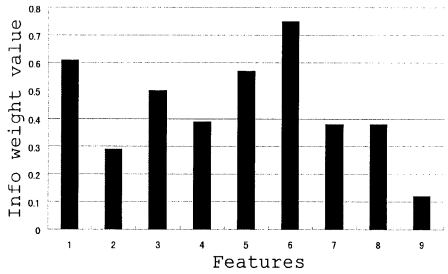

Fig. 4. Info weights in trial 1

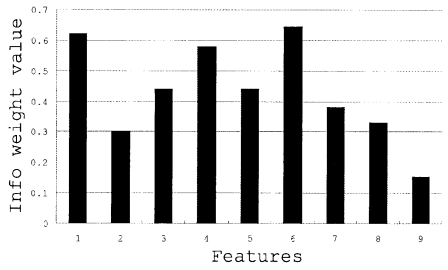

Fig. 5. Info weights in trial 2

Table 2. Results for *TF* classifiers on the breast cancer database

trial	classifier	rate	rejection
1	TFR_τ	100%	137
1	TF_τ	90.4%	–
1	TFR_T	97.4%	21
1	TF_T	96.8%	–
2	TFR_τ	100%	169
2	TF_τ	90.8%	–
2	TFR_T	96.7%	23
2	TF_T	95.4%	–

33 shortest prime tests are extracted to calculate an info vector. Fig. 5 shows a graphical interpretation of the info vector of a training set. The detected kernel is $\boldsymbol{\tau} = (101111000)$, which corresponds to 4 higher values of the info weights (since the 3th and 5th info weights have the same value, the length of $\boldsymbol{\tau}$ becomes 5). The performance of test feature classifiers are evaluated on two sets of tests, we use $\{\tau\}$ and the set T.

In Table 2, we show the recognition rates of the test feature classifiers. Although *TFR* rejects samples, it achieves a 100% recognition rate which means that the proposed kernels are real. This is a very important result and indicates that kernels can be detected from a training set even in difficult classification problem. Increasing the number of detected kernels will decrease the number of rejections. To improve the performance farther more powerful kernel detection algorithms need to be developed.

The number of rejections essentially decrease when we employ the set T in test feature classifiers. The performance of test feature classifiers on the set T is stable, remains high, and has small number of rejections..

Next, let us concentrate on the satimage database, which was generated from a multi-spectral satellite scan of landscape. This database is composed of six classes in 36 dimensions. Each feature is represented in eight bits, with 0 corresponding to black and 255 to white. There are 6435 patterns. We use the satimage_CR databases to evaluate test feature classifiers. The CR notation indicates that the database was preprocessed by a normalization routine in which each feature is centered and reduced to unit variance. The best estimate of the Bayes error rate of satimage_CR by a *k-NN* classifier using the Leave-one-out method is given in Table 3 [11].

Table 3. Best confusion matrix obtained by *k-NN* classifier on the satimage database using Leave-one-out method

class	1	2	3	4	5	6
1	98.1	0.2	1.1	0.1	0.5	0.0
2	0.0	96.5	0.1	0.7	2.0	0.7
3	0.5	0.1	93.4	4.6	0.0	1.4
4	0.0	0.8	13.7	70.6	0.8	14.1
5	3.1	0.8	0.1	0.8	89.7	5.5
6	0.0	0.1	1.9	7.3	2.0	88.7

In Woods et al. [12], where 50% of the data samples were used as training the best performance on the satimage_CR database was achieved by the nearest neighbor with almost 88% recognition rate. Experiments in Blayo et al. [11] (50% as training data) also showed that the nearest neighbor has the best performance with almost 90% recognition rate. The experiments with the same data [12] show that the satimage database is a quite difficult classification problem.

The satimage database is of multi-class, and the test concept presented here is for a two-class problem. We discuss the concept of a test for multi-classes [15], and since the purpose of this section is only to show a 100% recognition ability of test feature classifiers, we concentrate on only one two-class problem from the satimage database. As seen from Table 3, class 4 has the lowest recognition rate. This class represents damp grey soil, which appears difficult to discriminate from classes 3 and 6, which represent grey soil and very damp grey soil, respectively. Let us consider the problem of discriminating class 4 from the other classes. We apply test feature classifiers directly to the 36-dimensional real-valued satimage_CR database. The inequality R is defined as, $xR\,y$ if and only if $|x - y| > tr$, where tr is chosen such that it corresponds to the intensity value 4 in the original satimage database. This choice of threshold value was made by taking into account the fact that human eyes are almost insensitive to changes in intensity around 4.

In each trial the kernel detection algorithm is applied to the training samples. 50,000 short tests (with length no more $d = l + 1$, where $l = 15$ is minimal length of the detected tests) are extracted and the set T is formed. Then based on T we calculate an info vector, and construct a candidate kernel τ of length d, by taking d features with the highest info weights. A set T' is constructed from 300 randomly chosen prime tests. The performance of test feature classifiers are evaluated on the set T' and on the candidate kernel τ. In Table 4, we show recognition rates of the test feature classifiers.

Table 4. Results for *TF* classifiers on the satimage database

trial	classifier	rate	rejection
1	TFR_{τ}	100%	3128
1	TF_{τ}	90.4%	–
1	$TFR_{T'}$	90.6%	1305
1	$TF_{T'}$	91.1%	–
2	TFR_{τ}	100%	3145
2	TF_{τ}	90.4%	–
2	$TFR_{T'}$	90.1%	1485
2	$TF_{T'}$	90.2%	–

Again, a 100% of recognition rate is achieved by *TFR*, which means that the proposed kernels are real. The large number of rejections can be reduce by increasing the number of kernels, or by constructing a prototype training set. The error rate of *TF*, which has no rejections is much better than a Bayes error rate estimation obtained by *k-NN* with the Leave-one-out method.

5. Conclusions

Test feature classifiers are m-degree polynomials, and can be used for partitioning the n-dimensional feature space, $m \leq n$. Optimization methods, statistical, structural or metrical characteristics of patterns are not required. The method is desirable when statistical or structural information is not available. We discuss the generalization ability of the proposed classifiers. We show that test feature classifiers theoretically can achieve a 100% recognition rate on any test data. This is happens when a set of kernels is detected. In general, it is impossible to find kernels, but we proposed a heuristic algorithm to estimate them.

To test the performance of the classifiers, we apply them to the well-known breast cancer and satellite image databases. Experiments show that kernels can be detected from training sets, and confirm the ability of a 100% recognition rate, despite a large number of rejections. Possible ways to reduce this number are increasing the number of training samples or kernels, or constructing a prototype training set. Future research will be focus on the development of an efficient algorithm for construction of a prototype set.

(Manuscript received October 20, 12, revised November 7, 12)

References

(1) I. Chegis and S. Yablonsky, "Logical Methods for Controlling Electric Circuit Function", Proceedings of V. A. Steklov Inst. of Maths., Vol 51, (in Russian) 1958.

(2) Yu. Zhuravlov, A. Dmitriev and F. Krendelev, "Mathematical Principles of the classification of Objects and Scene", Discrete Analyze, Vol 7, 30-45, (in Russian) 1966.

(3) E. Djukova, "On Asymptotically Optimal Terminal Test Detection Algorithm ", Dokl. Akad. Nauk SSSR, 233, N. 4, 527-530, (in Russian) 1977.

(4) A. Kibkalo, T-algorithms that use short tests, Phd thesis, Moscow State University, (in Russian) 1988.

(5) S. V. Aleshin, Recognition of Dynamical Objects, Moscow University Press, (in Russian) 1996.

(6) D. Wilson,and T. Martinez, "Reduction Techniques for Instance-Based Learning Algorithms", Machine Learning, 38, Nov. 257-286, 2000.

(7) D.Aha, D. Kibler, and M. Albert, "Instance-Based Learning Algorithms", Machine Learning, 6, 37-66, 1991.

(8) J. Zhang, "Selecting Typical Instances in Instance-based Learning" Proceedings of the Ninth International Machine Learning Conference, 470–479, 1992.

(9) W. Wolberg, and O. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology", Proc. Natl. Acad. SCi. USA, Vol. 87, 9193-9196, 1990.

(10) O. Mangasarian, R. Setiono, and W.H. Wolberg: "Pattern recognition via linear programming: Theory and application to medical diagnosis", in: "Large-scale numerical optimization", T. Coleman and Y. Li, editors, SIAM Publications, Philadelphia, 22-30, 1990.

(11) F. Blayo et al., Enhanced Learning for Evolutive Neural Architecture, ESPRIT Basic Research Project, No. 6891, 1995.

(12) K. Woods, W. Kegelmeyer, and K. Bowyer, "Combination of Multiple Classifiers Using Local Accuracy Estimates", IEEE Trans. PAMI-19, 405-410, 1997.

(13) E. Gose, R. Johnsobaugh, S. Jost, Pattern Recognition and Image Analysis, Prentice Hall, 1996.

(14) V. Lashkia, S. Kaneko, and S. Aleshin, "Textual Region Location in Complex Images Using Test Feature Classifiers", Canadian J. Elect. Eng., Vol. 24, No. 2, 65-71, 1999.

(15) Itqon, S. Kaneko, S. Igarashii, V. Lashkia,"Extended Test Feature Classifier for Many-valued Patterns and its Experimental Evaluations", (in Japanese) IEEJ Trans., vol.120-C, no.11, 1762-1769, 2000.

(16) V. Lashkia et al., Soft Computing in Industrial Applications, Springer, 2000.

Vakhtang Lashkia (Non-member) received the M.S. and Dr. of Sci. degrees in computer science, from Moscow State University, in 1984 and 1988 respectively. He is now an associate professor of Okayama University of Science. Previous working places are: Moscow State University, Tbilisi University, and Hitotsubashi University. His research interests include machine learning, patter recognition and automata theory. He is a member of IEEE and ACM and IEICE.

Shun'ichi kaneko (Member) received the B.S. degree in precision engineering and the M.S. degree in information engineering from Hokkaido University, Japan, in 1978 and 1980, respectively, and then the Ph.D degree in systems engineering from the University of Tokyo, Japan, in 1990. He was an associate professor of the Department of Electronic Engineering since 1991 to 1996, in Tokyo University of Agriculture and Technology, Japan. He is an associate professor of Hokkaido University from 1996. He received the Best Paper Award in 1990, the Society Award in 1998, respectively, from JSPE. His research interest includes machine vision, image sensing and understanding, robust image registration. He is a member of JSPE, IEICE, IPSJ and IEEE.

Mitsuru Okura (Non-member) received the B.E. and M.E. degrees in mechanical engineering from Ehime University, Japan, in 1983 and 1985, respectively. He received the D.Sc. degree in system science from Okayama University of Science, Japan, in 1990. Since 1992 he has been a Lecture in the Department of Information and Computer Engineering, Faculty of Engineering, Okayama University of Science, engaged in education and research in image processing and pattern recognition. He is a member of IEICE, IPSJ and ITE.