An Evaluation of the Robustness of MTS for Imbalanced Data

Article *in* IEEE Transactions on Knowledge and Data Engineering · November 2007

DOI: 10.1109/TKDE.2007.190623 · Source: IEEE Xplore

Chao-Ton Su
National Tsing Hua University

Yu-Hsiang Hsiao
National Taipei University

An Evaluation of the Robustness of MTS for Imbalanced Data

Chao-Ton Su and Yu-Hsiang Hsiao

Abstract—In classification problems, the class imbalance problem will cause a bias on the training of classifiers and will result in the lower sensitivity of detecting the minority class examples. The Mahalanobis-Taguchi System (MTS) is a diagnostic and forecasting technique for multivariate data. MTS establishes a classifier by constructing a continuous measurement scale rather than directly learning from the training set. Therefore, it is expected that the construction of an MTS model will not be influenced by data distribution, and this property is helpful to overcome the class imbalance problem. To verify the robustness of MTS for imbalanced data, this study compares MTS with several popular classification techniques. The results indicate that MTS is the most robust technique to deal with the classification problem on imbalanced data. In addition, this study develops a "probabilistic thresholding method" to determine the classification threshold for MTS, and it obtains a good performance. Finally, MTS is employed to analyze the radio frequency (RF) inspection process of mobile phone manufacturing. The data collected from the RF inspection process is typically an imbalanced type. Implementation results show that the inspection attributes are significantly reduced and that the RF inspection process can also maintain high inspection accuracy.

Index Terms—Data mining, classification, class imbalance problem, imbalanced data, Mahalanobis-Taguchi System (MTS), threshold, mobile phone inspection.

1 Introduction

The classification problem is one of the main issues in data mining because the attempt is to extract a classifier that can be used to predict the classes of objects whose class labels are unknown. The binary classification problem, a subset of classification problems, is one in which the data are restricted to one of two groups. At present, these problems are often seen in product inspection, voice recognition, disease diagnosis, credit rating, and so on. Additionally, in order to execute a classification task efficiently, feature selection is usually merged into establishing a classifier. Through the employment of feature selection, a classifier can be established by fewer and more important variables. Now, a number of statistics, machine learning, and artificial intelligence techniques have been used to solve a binary classification problem and can accomplish the feature selection.

Classification techniques such as decision tree analysis, discriminate analysis, neural networks, and so on always assume that the training examples are evenly distributed among different classes. However, this is not always the situation in actual cases where one class might be represented by a large number of examples, whereas the other class, usually the more important class, is represented by only a few. For example, within a steady production process, especially in a six sigma process, few defected items can be collected as compared to the large number of nondefected ones. This situation is known as the class imbalance problem [1], [2], and it occurs frequently in many real-world applications such as fraud detection [3], product inspection, text classification [4], and medical diagnosis [5], [6]. When a classifier is learned from an imbalanced or skewed data set, it will cause bias and, then, the tendency is that the classifier will produce high predictive accuracy over the majority class, but will predict poorly over the minority class [7]. Furthermore, the examples in the minority class can be treated as noise and are ignored completely by the classifier.

Several researchers have studied at the data and algorithm levels to cope with the class imbalance problem. At the data level, the methods include many different forms of sampling [3], [8], [9]. The main concept of sampling methods is to balance class distribution by randomly replicating (oversampling) the minority class examples or eliminating (downsampling) the majority class examples or both. At the algorithm level, however, the methods include adjusting the cost matrices [5], moving the decision thresholds [10], [11], and so on. Adjusting the cost matrices assumes that the cost matrices are known for different types of errors and aims to improve the prediction performance by setting a high cost to the misclassification of a minority class example. On the other hand, moving the decision thresholds tries to adapt the thresholds to impose a bias on the minority class. However, a common problem for these methods is that they lack a rigorous and systematic treatment on imbalanced data and sometimes may cause problems [11]. For example, downsampling the data can potentially remove certain important information, whereas oversampling them may introduce noise. Moreover, adjusting the cost matrices or changing thresholds for each class lacks the systematic foundation in the same sense as the sampling method, which may end up with rules overfitting the training data [12]. Therefore, this study concentrated on

• *The authors are with the Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Engineering Building #1, 101, Sec. 2, Kuang-Fu Road, Hsinchu 30013, Taiwan. E-mail: ctsu@mx.nthu.edu.tw, adrian.iem88@nctu.edu.tw.*

Manuscript received 20 Apr. 2006; revised 26 Dec. 2006; accepted 8 May 2007; published online 7 June 2007.
For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number TKDE-0184-0406.
Digital Object Identifier no. 10.1109/TKDE.2007.190623.

finding a classification technique in which the training cannot be influenced by the imbalanced data.

The Mahalanobis-Taguchi System (MTS), developed by Taguchi, is a collection of methods proposed for a diagnostic and forecasting technique using multivariate data [13], [14], [15], [16], [17], [18] and has been used in various applications [19], [20], [21]. MTS combines Mahalanobis distance (MD) and Taguchi's robust engineering. MD is used to construct a multidimensional measurement scale and define a reference point of the scale with a set of observations from a reference group [13], [14], whereas Taguchi's robust engineering is applied to determine the important variables and then optimize the system. It must be noted that unlike most classification techniques, MTS establishes a classification model by constructing a continuous measurement scale using single class samples rather than directly learning from the whole training data set. This property seems useful in solving the class imbalance problems. On the other hand, determining an appropriate threshold is very important for MTS to carry out the classification process effectively. Taguchi and Jugulum suggested utilizing the loss function to determine a threshold [14]; however, this approach is not popular in practice because of the difficulty in exactly estimating the relative cost or loss in each case [15]. For this reason, this study used the Chebyshev's theorem to propose a "probabilistic thresholding method" (PTM) for MTS.

This study was carried out in order to investigate whether or not MTS has a more robust classification ability than other classification techniques when facing class imbalance problems. Meanwhile, we also evaluated the performance of the PTM. Finally, a real case about improving the mobile phone radio frequency (RF) inspection process was employed to illustrate the practicality of MTS. Through the application of MTS, the redundant inspection attributes were efficiently detected and removed without losing inspection accuracy.

2 Mahalanobis-Taguchi System

MTS was developed by Taguchi as a diagnostic and forecasting technique using multivariate data. In a typical multidimensional system, there is always more than one variable that provide information that can be used to make a decision. However, one can make wrong decisions if each variable is looked at separately without considering its correlation structure with other variables. The MD introduced by Mahalanobis in 1936 takes the correlation structure of a system into account. In MTS, MD is scaled by dividing the original one by the number of variables and then employed to construct a multidimensional measurement scale. MD is calculated as follows:

$$MD = 1\Big/k \cdot Z^T \cdot C^{-1} \cdot Z, \tag{1}$$

where

- k is the number of variables,
- Z is the standardized vector of the example, and
- C is the correlation matrix of the reference group.

Taguchi's robust engineering was proposed by Genichi Taguchi in the 1950s, and it aims to improve the engineering quality that can be measured in terms of deviations from the ideal performance. MD can be regarded as a kind of engineering quality because it describes the degree of abnormality of observations from the known reference group. MTS combines MD with Taguchi's robust engineering and is used for optimizing the multidimensional systems.

To implement MTS, the first step is to identify a "reference" or "normal" group, which is used to construct the Mahalanobis space (MS). MS can be regarded as a database for the normal group consisting of its mean vector, standard deviation vector, and correlation matrix [14]. In general, the examples in the normal group should be similar and have common characteristics. Take medical diagnosis for example, MS is constructed only using the people who are healthy. The mean point and the average MD of the normal group serve as the reference point and the base of the measurement scale. To validate the scale, different known "abnormal" examples must be checked. The MDs of abnormal examples are also computed using the information contained in the MS. If the scale is good, the MDs corresponding to these examples should match with the judgment. Otherwise, it implies that the MS cannot suitably represent the real normal condition and is necessary to be reconstructed. A good MS is very important for MTS. However, it sometimes may be difficult to select an MS, especially when we deal with historical data, in which the "normal group" cannot be well identified due to the disorganization of the data. In such situations, we can first generate the MS using all normal examples and calculate their MDs. Then, we discard the examples having higher MDs and recalculate MDs with a new MS constructed using the remaining normal examples. This process is repeated until we generate a suitable MS for conducting a multivariate diagnosis [14].

In the next phase, orthogonal arrays (OAs) and signal-to-noise (SN) ratios are used to screen the important variables. Applying OAs, each variable is assigned to one column and set with two levels: using and not using this variable. The SN ratio, the larger-the-better SN ratio is frequently suggested [13], [14], obtained from the abnormal MDs is used as the response for each run of OA. The importance of each variable is evaluated by calculating the "effect gain." If the gain corresponding to a variable is positive, the variable may be considered as worth keeping; otherwise, it should be removed. Finally, a "reduced model measurement scale" is established using the important variables. Then, an appropriate threshold to discriminate between the normal group and the abnormal examples are determined for future diagnosis. Fig. 1 shows the procedure of implementing MTS.

MTS is different from classical multivariate methods in the following ways [13], [14]. First, the methods used in MTS are data analytic rather than being on probability-based inference. That is, MTS does not require any assumptions on the distribution of input variables. Second, the MD in MTS is suitably scaled and used as a measure of severity of various conditions. MTS can be used not only to

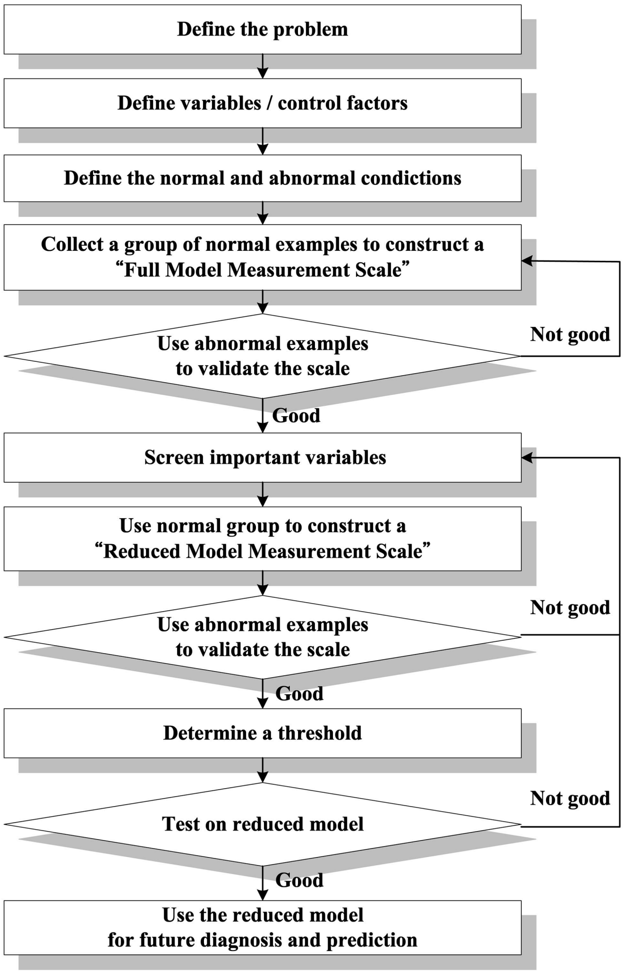

Fig. 1. The implementing procedures of MTS.

classify observations into two different groups (that is, normal and abnormal) but also to measure the degree of abnormality of an observation. Third, every example outside the normal space (that is, abnormal example) is regarded as unique and does not constitute a separate population.

3 Proposed Probabilistic Thresholding Method

In MTS, the MD distributions of normal and abnormal examples usually overlap. An effective threshold can enhance the diagnostic and forecasting ability of MTS. However, how to find an appropriate threshold to effectively distinguish the normal and abnormal examples is an important issue. Taguchi and Jugulum suggested using the "quadratic loss function" to determine a threshold [14], yet this method is impractical because of the difficulty in estimating the relative cost or loss in each case [15]. Instead, real applications always use the "exhaustive search method" (ESM) to search the threshold that results in the highest classification accuracy on the training set. However, the ESM is time consuming, and the determined threshold may cause the MTS model to overfit the training set and lower the classification reliability. This study employed Chebyshev's theorem to develop a PTM. Chebyshev's theorem is useful to estimate the probability of getting a value that deviates from the mean by less than some degree of standard deviation, especially when the probability distribution of the data set is unknown. Our PTM applies this probability property to compute a threshold for MTS.

According to Chebyshev's theorem, it can be easily proved that the following inequality is valid:

$$P(\mu_X + r \cdot \sigma_X \leq X) \leq \frac{1}{r^2}, \tag{2}$$

where μ_X and σ_X are the mean and the standard deviation, respectively, of the random variable X. The probability of getting a value that is larger than μ_X by at least $r\sigma_X$ is at most $\frac{1}{r^2}$. On the basis of MTS theory, only normal examples can constitute a population, whereas each abnormal example is unique. Therefore, we regard the normal MD as a random variable X_{md} that is generated from an unknown probability distribution, and μ_{md} and σ_{md} are its mean and standard deviation. Let T denote our proposed threshold that discriminates between normal and abnormal examples, and the probability of false alarm is denoted by $P(T \leq X_{md})$.

Now, we used the information provided by the training set and (2) to develop a threshold. The steps are detailed as follows:

Step 1. Remove the outliers, the examples with MDs out of the three standard deviations, on the left tail of the MD distribution of the abnormal examples

Step 2. Among the normal group, compute the percentage of the examples whose MDs are smaller than the minimum MD of the remainder abnormal examples. The percentage is denoted by ω. Because some abnormal MDs on the left tail have been removed, a small parameter λ ($\omega > \lambda > 0$), 0.05 is typical, is introduced to adjust the percentage ω to reflect the real situation of the nonoverlap of the normal group. Thus, under the MD boundary covering $\omega - \lambda$ of the normal distribution, it is confident that there will be a perfect prediction on abnormal examples and a corresponding maximum false alarm percentage $1 - (\omega - \lambda)$ on the normal group.

Step 3. Set the upper bound of the false alarm probability $P(T \leq X_{md})$ equal to empirical percentage $1 - (\omega - \lambda)$:

$$P(T \leq X_{md}) \leq 1 - (\omega - \lambda). \tag{3}$$

Step 4. Screen (3); if the μ_{md} and the σ_{md} of random variable X_{md} are known and the upper bound probability is also known, we can apply Chebyshev's theorem (2) to obtain the value of the corresponding lower bound of X_{md}, T. Let $1 - (\omega - \lambda) = \frac{1}{r^2}$; we can compute $r = \sqrt{\frac{1}{1+\lambda-\omega}}$ and, thus, $T = \mu_{md} + \sqrt{\frac{1}{1+\lambda-\omega}} \cdot \sigma_{md}$.

Step 5. For the unknown of μ_{md} and σ_{md}, use the average and standard deviation of MDs of the normal group, $\overline{md}$ and s_{md}, to be their unbiased estimators, respectively. Thus, the threshold can be calculated with the following formula:

TABLE 1
Summary Descriptions of the Data Sets

Data Set	Class Maj. / Min.	# Var.	Exp.	*f*-ratio	Training Set		Test Set	
					# Neg.	# Pos.	# Neg.	# Pos.
Breast Wisconsin	Normal / Abnormal	9	1	3.52	66	33	22	11
			2	3.55	165	33		
			3	3.40	363	33		
Letter	A / Remainder	16	1	1.60	90	45	30	15
			2	1.58	405	45		
			3	1.78	675	45		
Heart Disease	Normal / Abnormal	13	1	2.63	60	30	30	15
			2	2.67	180	30		
			3	2.66	300	30		
Shuttle	Remainder / Class 5	9	-	11.94	50000	45	200	200
Covtype	Class 1 / Class 4	12	-	41.46	65516	20	300	300

$$T = \overline{md} + \sqrt{\frac{1}{1+\lambda-\omega}} \cdot s_{md}, \tag{4}$$

where

- $\overline{md}$ is the average of the MDs of the normal group,
- s_{md} is the standard deviation of the MDs of the normal group,
- λ is a small parameter, and
- ω is the percentage of the normal examples whose MDs are smaller than the minimum MD of the remainder abnormal examples.

4 Robustness Evaluation

In order to evaluate classifiers on the imbalanced data sets, the medical community and, increasingly, the machine learning community use two metrics, that is, sensitivity and specificity. The sensitivity metric is defined as the accuracy on the positive examples (true positives/(true positives + false negatives)), whereas specificity is the accuracy on the negative examples (true negatives/(true negatives + false positives)). The abovementioned "negative" is taken as the majority class, whereas "positive" is the minority class. The g-means metric suggested by Kubat and Matwin [22] has been used by several researchers for evaluating classifiers on the imbalanced data sets [9], [23], [24]. This metric takes sensitivity and specificity into account simultaneously, which is defined as

$$g = \sqrt{Sensitivity \cdot Specificity}. \tag{5}$$

Relative sensitivity (RS) was proposed in this study as follows:

$$RS = \frac{Sensitivity}{Specificity}. \tag{6}$$

The RS can be used to judge whether or not a classifier has the balanced ability to predict the positive and the negative examples. If the RS is much lower or higher than 1, it indicates that the classifier has some bias on classifying. This study simultaneously used the g-means and the RS to be the two main metrics to evaluate the performance of classifiers. Besides, the sensitivity and specificity will also be listed separately to give the reader an even better idea of the performance of each classifier.

4.1 MTS versus Prevalent Classification Techniques

In this section, we compared the performance of MTS with other popular classification techniques such as the stepwise discriminate analysis (SDA), decision tree analysis (C4.5), back-propagation neural network (BPN), and support vector machines (SVMs). In the meantime, we used the proposed PTM and the ESM, respectively, to determine the threshold for MTS (that is, MTS(PTM) and MTS(ESM) models), and then, we compared their classification results. For the purposes of investigating how the class imbalance problem impacts on the training of a classifier and checking if MTS can have the robust ability to overcome it, four different UCI data sets and one Statlog data set were utilized. Table 1 shows the characteristics of these five data sets. For each of the top three data sets, three training sets with different degrees of class imbalance were designed for training (that is, experiments 1, 2, and 3), and the same test set was used for evaluating the trained classifiers. The distribution of the training set in experiment 1 was just imbalanced slightly, whereas from experiment 1 to experiment 3, the imbalance level was more and more significant. Through comparing the test results corresponding to the experiments 1, 2, and 3, we were able to trace if the training of a classifier is affected by the different degrees of class imbalance of the training set. The bottom two data sets were much more imbalanced than the top three ones, and the negative-to-positive ratios were more than 1,000:1. In addition to the different class imbalance levels, these problems involved widespread data sizes from dozens to ten thousands and various complexities that were measured by the maximum Fisher's Discriminant Ratio over all the feature dimensions (f-ratio) [25]. A smaller f-ratio means that the two classes have more overlap regions and, thus, more difficult to deal with.

Table 2 shows the test results on the Breast-Wisconsin data set. The results indicated that MTS(PTM) performed slightly worse than SVMs but better than the other methods. In experiment 1, the two classes of the training set were

TABLE 2
Comparison of Test Results on the Breast-Wisconsin Data Set

Data Set	Exp.	Index	MTS(PTM)	MTS(ESM)	SDA	C4.5	BPN	SVMs
Breast Wisconsin	1	Sensitivity	100.00	100.00	100.00	90.91	100.00	100.00
		Specificity	95.45	100.00	100.00	100.00	100.00	100.00
		g-means	97.70	100.00	100.00	95.35	100.00	100.00
		RS	1.05	1.00	1.00	0.91	1.00	1.00
	2	Sensitivity	90.91	90.91	81.82	63.64	77.78	90.91
		Specificity	95.45	95.45	100.00	95.49	91.67	100.00
		g-means	93.15	93.15	90.45	77.96	84.44	95.35
		RS	0.95	0.95	0.82	0.67	0.85	0.91
	3	Sensitivity	100.00	90.91	81.82	45.45	81.82	100.00
		Specificity	100.00	100.00	100.00	100.00	100.00	100.00
		g-means	100.00	95.35	90.45	67.42	90.45	100.00
		RS	1.00	0.91	0.82	0.45	0.82	1.00

TABLE 3
Comparison of Test Results on the Letter Data Set

Data Set	Exp.	Index	MTS(PTM)	MTS(ESM)	SDA	C4.5	BPN	SVMs
Letter	1	Sensitivity	100.00	100.00	100.00	86.67	93.33	100.00
		Specificity	100.00	100.00	93.33	80.00	90.00	93.33
		g-means	100.00	100.00	96.61	83.27	91.65	96.61
		RS	1.00	1.00	1.07	1.08	1.04	1.07
	2	Sensitivity	93.33	80.00%	86.67	80.00	93.33	93.33
		Specificity	96.67	100.00%	100.00	100.00	100.00	100.00
		g-means	94.98	89.44%	93.10	89.44	96.61	96.61
		RS	0.97	0.80	0.87	0.80	0.93	0.93
	3	Sensitivity	93.33	93.33%	53.33	60.00	86.67	80.00
		Specificity	93.33	96.67%	96.67	100.00	100.00	100.00
		g-means	93.33	94.98%	71.80	77.46	93.10	89.44
		RS	1.00	0.97	0.57	0.60	0.87	0.80

TABLE 4
Comparison of Test Results on the Heart-Disease Data Set

Data Set	Exp.	Index	MTS(PTM)	MTS(ESM)	SDA	C4.5	BPN	SVMs
Heart Disease	1	Sensitivity	86.67	86.67	73.77	46.67	73.33	40.00
		Specificity	83.33	83.33	86.67	96.67	90.00	90.00
		g-means	84.98	84.98	79.96	67.17	81.24	60.00
		RS	1.04	1.04	0.85	0.48	0.81	0.44
	2	Sensitivity	93.33	80.00	66.67	20.00	46.67	40.00
		Specificity	90.00	93.33	86.67	96.67	96.67	96.67
		g-means	91.65	86.41	76.02	43.97	67.17	62.18
		RS	1.04	0.86	0.77	0.21	0.48	0.41
	3	Sensitivity	93.33	53.33	53.33	13.33	33.33	46.67
		Specificity	93.33	96.67	86.67	96.67	96.67	96.67
		g-means	93.33	71.80	67.99	35.90	56.76	67.17
		RS	1.00	0.55	0.62	0.14	0.34	0.48

slightly imbalanced, and all the six methods had good performance on the sensitivity and specificity metrics. However, when the degree of class imbalance of the training set was increased, as done in experiments 2 and 3, the sensitivity of SDA, C4.5, and BPN significantly decreased although the specificity was still at a high level. By contrast, MTS(PTM) and SVMs could have high sensitivity and specificity even if the training sets were much imbalanced, and this resulted in good g-means metrics on the three experiments. The RSs of MTS(PTM) and SVMs on the three experiments were all near to 1, whereas the RS of SDA, C4.5, and BPN decreased as the degrees of class imbalance increased. It appears that the ability of SDA, C4.5, and BPN classifiers to predict the positive examples decreases as the proportion of the positive instances in the training set decreases. In the Letter data set, the test results are in Table 3. It is clear that only MTS(PTM) could achieve high g-means and keep RSs to be around 1. The Heart-Disease data set has a significant overlap between its two classes. Thus, to determine an appropriate threshold to correctly differentiate the two classes seems more important for MTS. Table 4 shows the test results on the Heart-Disease data set. The performance of MTS(PTM) was much better than that of the other five methods. Additionally, we also found that the specificity of SDA, C4.5, BPN, and SVMs were much higher than the sensitivity, especially in experiment 3. In experiment 3, the C4.5 classifier almost lost the ability to predict the positive examples. Table 5 shows the test results on the Shuttle and Covtype data sets. Under a negative-to-positive ratio more than 1,000:1, MTS(PTM) still performed robustly on g-means and RS and was better than the other methods.

TABLE 5
Comparison of Test Results on the Shuttle and Covtype Data Sets

Data Set	Index	MTS(PTM)	MTS(ESM)	SDA	C4.5	BPN	SVMs
Shuttle	Sensitivity	100.00	97.78	100.00	0.00	0.00	100.00
	Specificity	100.00	100.00	98.00	100.00	100.00	100.00
	g-means	100.00	98.88	98.99	0.00	0.00	100.00
	RS	1.00	0.98	1.02	0.00	0.00	1.00
Covtype	Sensitivity	100.00	100.00	100.00	0.00	0.00	37.67
	Specificity	100.00	100.00	99.33	100.00	100.00	100.00
	g-means	100.00	100.00	99.66	0.00	0.00	61.38
	RS	1.00	1.00	1.01	0.00	0.00	0.38

TABLE 6
Mean and Standard Deviation of g-Means Prediction Accuracy on UCI Data Sets

Dataset	f-ratio	# Var.	# Neg.	# Pos.	SVMs	SMOTE	ACT	KBA	MTS(PTM)
Car	1.01	6	1659	69	99.0 ± 2.2	99.0 ± 2.3	99.9 ± 0.2	99.9 ± 0.2	99.9 ± 0.2
Yeast	1.24	8	1433	51	59.0 ± 12.1	69.9 ± 10	78.5 ± 4.5	82.2 ± 7.1	77.6 ± 2.8
Abalone	0.52	8	4145	32	0.0 ± 0.0	0.0 ± 0.0	51.9 ± 7.6	57.8 ± 5.4	69.4 ± 12.6

We can see that the SVM classifier could not predict with good accuracy in all cases, and the BPN and C4.5 classifiers entirely lost their ability to predict the minority class.

4.2 MTS versus Modified SVMs

SVMs have been extensively studied and successfully applied to many domains such as image processing [26], text classification [27], and so on. However, when facing the class imbalance problem, the classification performance of SVMs drops significantly [23], [24], [28], [29]. Up to the present, a lot of research has been proposed for SVMs to deal with imbalanced data sets, such as the Synthetic Minority Oversampling Technique (SMOTE) [4] at the data level, and Adaptive Conformal Transformation (ACT) [29] and Kernel Boundary Alignment (KBA) [23] at the algorithm level. The experimental results of these techniques or algorithms showed their good ability to overcome the class imbalance problem [23], [29]. In order to justify the superiority of MTS(PTM) for imbalanced data classification, we compared MTS(PTM) with regular SVMs, SMOTE, ACT, and KBA. In this comparison, we employed three of the most imbalanced UCI data sets used in [23], and the measure of complexity (f-ratio) of each problem was appended to Table 6. In each run, the training and the test subsets were similarly generated in the ratio 6:1, and the setting of each algorithm was identical to that in [23]. Based on the results obtained from [23] and our experiments, the means and standard deviations of the classification g-means are both reported in Table 6. It was found that MTS(PTM) achieved a good prediction accuracy. MTS(PTM) tied for the highest accuracy in the Car data set and was slightly lower than ACT and KBA in the Yeast data set. However, in the Abalone data set with most imbalanced ratio, MTS(PTM) had the highest accuracy among all methods and, meanwhile, we can observe that both regular SVMs and SMOTE performed poorly.

4.3 Discussion

In Section 4.1, we in total employed 11 experiments with different imbalance levels, sample sizes, and classification complexities to evaluate the performance of different classification techniques. The results indicated that using the PTM to determine a threshold can help MTS to avoid overfitting the training sets and to attain better classification ability than using ESM. The experimental results also revealed that the classification techniques, that is, SDA, C4.5, BPN, and SVMs, were influenced by the class imbalance problem much more easily than MTS regardless of the data sizes or complexities of the problems. Given the observance made on the g-means metric, MTS could almost make the highest value in the 11 experiments. Besides, the RSs of MTS steadily kept close to 1, whereas those of the other methods decreased with the increase in the degrees of class imbalance of the training sets. These performances mean that the ratio of the classes in the training set exercises influence over the learning procedures of the SDA, C4.5, BPN, and SVMs classifiers. On the contrary, MTS is robust not only when facing the class imbalance problem but also under the various data sizes and different problem complexities. In Section 4.2, we compared MTS with several modified SVMs devised for the class imbalance problem. The comparison results verified the good ability of MTS to deal with imbalanced data and conformed that MTS is not inferior to the modified SVMs.

Besides, from the viewpoint of implementation, there are no parameters needed to be optimized in executing MTS, whereas other techniques such as BPN and SVMs always spend much time on determining the best combination of parameters. Moreover, it is unnecessary for MTS to do any adjustment or modification at the data or algorithm level in order to fit imbalanced data learning. On the other hand, correctly predicting the positive examples is always more important in real applications because they affect the performance of a system. For example, if a production inspection system cannot correctly and immediately detect

TABLE 7
Attribute Value, Mean, and SD of the Normal Group

Sample \ Attribute	C_1	C_2	C_3	C_4	...	C_{61}	C_{62}
1	32.220	1.117	0	0	...	1	0
2	32.191	1.076	0	0.014	...	1	1
3	32.411	1.555	0	0.014	...	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
270	32.204	1.094	0	0.029	...	1	1
$\overline{x}_i$	32.22	1.467	0.003	0.021	...	0.652	0.511
s_i	0.082	0.24	0.019	0.017	...	0.493	0.523

TABLE 8
Standardized Attribute Value and MD of the Normal Group

Sample \ Attribute	C_1	C_2	C_3	C_4	...	C_{61}	C_{62}	MD
1	0	-1.458	-0.158	-1.235	...	0.706	-0.977	1.07470
2	-0.354	-1.629	-0.158	-0.412	...	0.706	0.935	0.91999
3	2.329	0.367	-0.158	-0.412	...	-1.323	0.935	1.09532
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
270	-0.195	-1.554	-0.158	0.471	...	0.706	0.935	0.99650

the defective products, it will entail a huge cost later. Therefore, a classification technique should have the ability to overcome the class imbalance problem. It is apparent that the ability of MTS to predict both of the classes is robust and is not affected by the distribution of the training set even though it is much imbalanced. This property is useful for real applications.

5 Case Study

5.1 Case Description

The case studied here comes from a mobile phone manufacturer located in Taiwan [30]. The RF functional inspection aims to inspect if a dual-band mobile phone receive/transmit signal satisfies the enabled transmission interval (ETI) protocol on different channels and power levels. In order to ensure the quality of communication of mobile phones, the manufacturer added many extra inspection items to the RF functional inspection process, such as several different frequency channels and power levels. However, the added inspections caused an increase in the required operation time and made the RF functional inspection process the bottleneck of the entire mobile phone manufacturing procedure.

The RF functional inspection process includes 62 inspection attributes, and they were labeled as C_1 to C_{62}. The purpose of this case study was to remove the redundant RF functional inspection attributes by using MTS to reduce the production costs, to shorten the time to market, and to enhance market competition.

5.2 Implementation

In this case study, the products passing the 62 RF functional inspection attributes were defined as the normal condition. The data were randomly sampled and split into training and test sets. The training set used to construct a measurement scale contained 270 normal and 30 abnormal examples, whereas the test set used to demonstrate the capability of the scale contained 90 normal and 10 abnormal examples.

5.2.1 Phase 1: Construct a "Full Model Measurement Scale" with MS as the Reference

The 270 normal examples in the training set were designed as the reference (normal) group. At first, the attribute values, means, and SDs of the normal group were collected, calculated, and depicted in Table 7 and, then, the standardized values of the 62 attributes were computed and shown in Table 8.

Next, we used the standardized attribute values in Table 8 to compute the inverse of the correlation matrix of the normal group in Table 9. Finally, we applied (1) to calculate the MDs of the normal group. For example, the MD of the first example of the normal group was calculated as follows:

$$MD_1 = \frac{1}{62}[0 \quad -1.458 \quad \cdots \quad -0.977]_{1\times 62} \begin{bmatrix} 16.69 & -1.35 & \cdots & 0.008 \\ -1.35 & 3.013 & \cdots & 0.214 \\ \vdots & \vdots & \ddots & -0.069 \\ 0.008 & 0.214 & \cdots & 1.457 \end{bmatrix}_{62\times 62} \begin{bmatrix} 0 \\ -1.458 \\ \vdots \\ -0.977 \end{bmatrix}_{62\times 1} = 1.0747.$$

The MD of each example of the normal group is shown in Table 8. These MDs defined an MS, and this space was taken as a reference point for the measurement scale.

5.2.2 Phase 2: Validate the Measurement Scale

The MDs corresponding to the 30 abnormal examples in the training set were also calculated using (1) to validate the accuracy of the scale. If the measurement scale constructed in phase 1 is good, the MDs of the abnormal examples will be larger than that of the normal group. The MD distributions of the normal group and the 30 abnormal examples were

TABLE 9
Inverse of the Correlation Matrix of the Normal Group

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	...	C_{58}	C_{59}	C_{60}	C_{61}	C_{62}
C_1	16.69	-1.35	0.254	-0.033	0.182	0.008	-0.565	...	0.038	0.224	-0.311	-0.078	0.008
C_2	-1.35	3.013	-0.073	0.107	0.109	0.035	-0.203	...	0.032	-0.123	0.041	-0.021	0.214
C_3	0.254	-0.073	1.17	0.107	-0.096	0.07	0.006	...	5E-04	0.031	0.023	-0.057	-0.069
C_4	-0.033	0.107	0.107	1.301	-0.037	-0.158	-0.036	...	0.029	0.289	0.035	0.193	0.024
C_5	0.182	0.109	-0.096	-0.037	1.734	-0.206	-0.146	...	0.12	-0.099	-0.133	-0.054	-0.022
C_6	0.008	0.035	0.07	-0.158	-0.206	1.793	0.137	...	-0.054	-0.267	-0.191	-0.04	-0.160
C_7	-0.565	-0.203	0.006	-0.036	-0.146	0.137	1.575	...	0.169	0.048	-0.023	-0.235	0.075
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
C_{58}	0.038	0.032	5E-04	0.029	0.12	-0.054	0.169	...	1.565	0.238	-0.1	0.016	-0.474
C_{59}	0.224	-0.123	0.031	0.289	-0.099	-0.267	0.048	...	0.238	1.539	-0.1	0.25	-0.004
C_{60}	-0.311	0.041	0.023	0.035	-0.133	-0.191	-0.023	...	-0.1	-0.1	1.35	0.049	0.089
C_{61}	-0.078	-0.021	-0.057	0.193	-0.054	-0.04	-0.235	...	0.016	0.25	0.049	1.444	-0.018
C_{62}	0.008	0.214	-0.069	0.024	-0.022	-0.16	0.075	...	-0.474	-0.004	0.089	-0.018	1.457

depicted in Fig. 2. It is obvious that the MDs of the abnormal samples are indeed larger than that of normal groups, and it appears that the measurement scale is effective.

5.2.3 Phase 3: Identify Important Variables (Feature Selection Phase)

In this phase, we regarded the 62 attributes as the control factors. Each factor was set into two levels, that is, level 1 was inclusive of the factor, and level 2 was exclusive of the factor. The 62 factors were allocated to the first 62 columns of an $L_{64}(2^{63})$ array. For each run of the OA, we used the factors with level 1 to construct an MS and, then, the MDs corresponding to the 30 abnormal examples of the training set were computed on the basis of the MS. Using the MDs corresponding to the 30 abnormal examples, the larger-the-better SN ratio was calculated for each run. The allocation of the factors in the OA and the SN ratios are shown in Table 10. Take run 1, for instance, the SN ratio was calculated as follows:

$$\eta_1 = -10 \cdot \log_{10}\left[\frac{1}{30} \cdot \left(\frac{1}{7.98674} + \cdots + \frac{1}{564,802}\right)\right] = 18.02903.$$

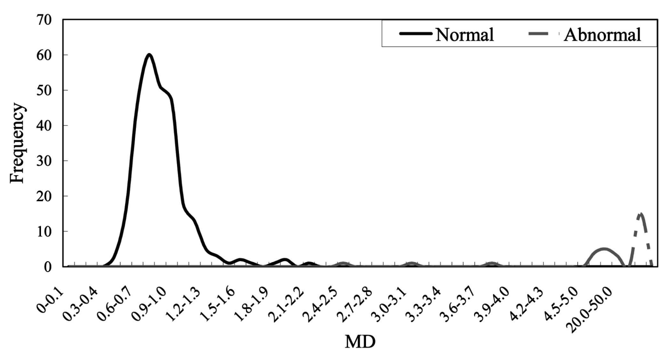

Fig. 2. MD distributions of the RF functional inspection training set (full model).

After obtaining the SN ratio of each run, we computed the effect gain of each attribute and plotted them into a graph, as shown in Fig. 3. Take attribute C_1 as an example, the effect gain was calculated as follows:

$$\begin{aligned}\overline{SN_1^+} &= \frac{1}{32} \cdot (18.02903 + 7.608896 + \cdots + 11.64836) \\ &= 12.91363 \\ \overline{SN_1^-} &= \frac{1}{32} \cdot (14.42624 + 12.68887 + \cdots + 13.35158) \\ &= 12.75443 \\ Gain_1 &= \overline{SN_1^+} - \overline{SN_1^-} = 12.91363 - 12.75443 = 0.1592.\end{aligned}$$

If the gain is positive, the attribute can be considered as worth keeping; if it is negative, the attribute should be removed.

5.2.4 Phase 4: Future Predict with Important Variables

The number of the inspection attributes was reduced from 62 to 36, 31, 27, 24, 17, 14, 10, 8, 7, and 6, respectively, according to the different gains (> 0, > 0.1, > 0.2, > 0.3, > 0.4, > 0.5, > 0.6, > 0.7, > 0.8, and > 0.9).

Take the case of gain > 0.5, the remainder attributes are C_7, C_{10}, C_{26}, C_{29}, C_{32}, C_{34}, C_{36}, C_{40}, C_{42}, C_{44}, C_{50}, C_{53}, C_{55}, and C_{60}, and we used the normal group with these 14 attributes to develop a reduced model measurement scale. Similarly, the 30 abnormal examples were then employed to validate the scale. Fig. 4 shows the MD distributions under the reduced model, and it indicates that the new scale is good. After making sure that the reduced model measurement scale was efficient, we applied the PTM to determine a threshold. The threshold was calculated by (4) as follows:

$$\begin{aligned}\mathrm{T}_{(14)} &= \overline{md} + \sqrt{\frac{1}{1+\lambda-\omega}} \cdot s_{md} \\ &= 0.982 + \sqrt{\frac{1}{1+0.05-1}} \cdot 0.393 = 2.724.\end{aligned}$$

TABLE 10
Factors Allocation and SN Ratios

Run	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	…	C_{60}	C_{61}	C_{62}		MD_1	…	MD_{30}	SN ratio
	1	2	3	4	5	6	7	8	9	10	…	60	61	62	63				η（dB）
1	1	1	1	1	1	1	1	1	1	1	...	1	1	1	1	7.98674	…	564802	18.02903
2	1	1	1	1	1	1	1	1	1	1	...	2	2	2	2	12.9013	…	978550	7.608896
3	1	1	1	1	1	1	1	1	1	1	...	2	2	2	2	1.50223	…	339314	12.57003
4	1	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1.60727	…	505956	14.05007
5	1	1	1	1	1	1	1	2	2	2	...	2	2	2	2	1.27893	…	752904	12.57109
6	1	1	1	1	1	1	1	2	2	2	...	1	1	1	1	1.60313	…	622553	11.40608
7	1	1	1	1	1	1	1	2	2	2	...	1	1	1	1	12.6725	…	39027.3	12.94864
8	1	1	1	1	1	1	1	2	2	2	...	2	2	2	2	11.873	…	188601	16.63614
9	1	1	1	2	2	2	2	1	1	1	...	2	2	2	2	1.42267	…	978514	12.81961
10	1	1	1	2	2	2	2	1	1	1	...	1	1	1	1	1.68229	…	117000	13.28187
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	2	2	1	2	1	1	2	1	2	2	...	2	1	1	2	1.01236	..	489627	10.97286
61	2	2	1	2	1	1	2	2	1	1	...	1	2	2	1	1.55941	…	778281	12.66432
62	2	2	1	2	1	1	2	2	1	1	...	2	1	1	2	1.34182	…	602506	13.4991
63	2	2	1	2	1	1	2	2	1	1	...	2	1	1	2	9.97475	…	38333	11.44389
64	2	2	1	2	1	1	2	2	1	1	...	1	2	2	1	9.29266	…	194704	13.35158

For this reduced model with 14 attributes, using 2.724 to be the threshold resulted in 100 percent classification accuracy on the training set, which can be observed in Fig. 4. Finally, to verify the classification capability of the reduced model, the test set was utilized. The MDs of the examples of the test set were also computed by (1). Fig. 5 shows the MD distributions of the test set. By continuously applying 2.724 to be the threshold, the classification accuracy was 100 percent.

In the same way as described above, we can also obtain the analysis results of the other reduced models according to the different effect gains. Table 11 summarizes the remainder numbers of attributes, the threshold calculated using PTM, and the training and test classification accuracy of each reduced model. The reduced models with the gain larger than 0.2, 0.3, 0.4, and 0.5 all had perfect training and test classification accuracy. Clearly, we should adopt the reduced model with 14 attributes to be the final outcome of this case because it had the best performance with the least number of attributes. That is, we can use only the 14 test attributes (C_7, C_{10}, C_{26}, C_{29}, C_{32}, C_{34}, C_{36}, C_{40}, C_{42}, C_{44}, C_{50}, C_{53}, C_{55}, and C_{60}) instead of the original 62 attributes for the mobile phone RF functional test process.

5.3 The Benefit

This case study drew support from MTS to analyze the RF functional inspection process of dual-band mobile phone manufacturing. The result indicated that the number of the RF functional test attributes was significantly reduced from 62 to 14 without losing classification accuracy. In virtue of attribute reduction, the operation time of the RF functional inspection process was reduced from 190 to 110 seconds. Because the bottleneck of the entire mobile phone manufacturing procedure was broken, the throughput increased from 18 to 32 in one hour, that is, the production efficiency is close to the double of the past. Additionally, this led to the number of the RF functional inspection machine reducing from 8 to 5, which saved the manufacturer a direct cost of five million New Taiwan (NT) dollars per year. With MTS, the redundant inspection attributes were removed so that the required operation time was cut down, and this resulted in a smoother flow of production and enhanced market competitiveness.

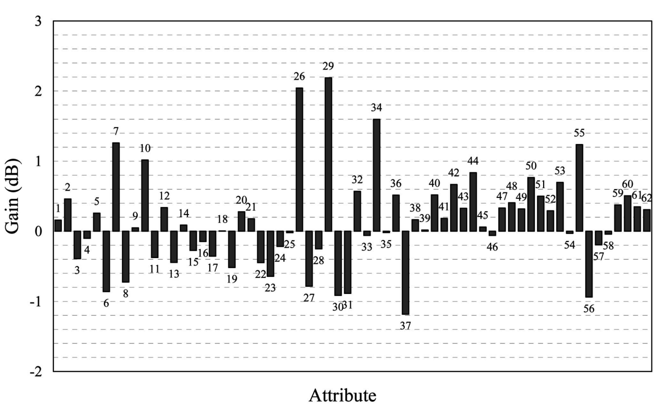

Fig. 3. The effect gains of the RF functional inspection attributes.

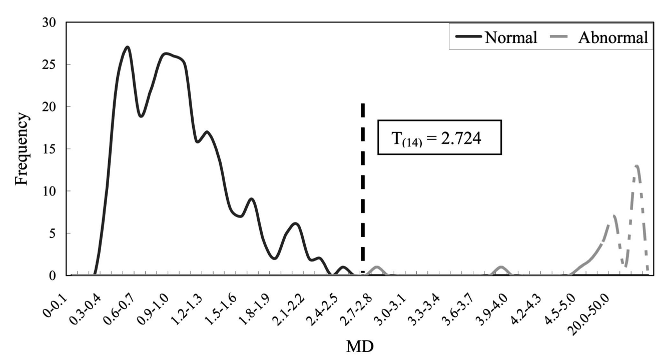

Fig. 4. MD distributions of the RF functional inspection training set (reduced model with 14 attributes).

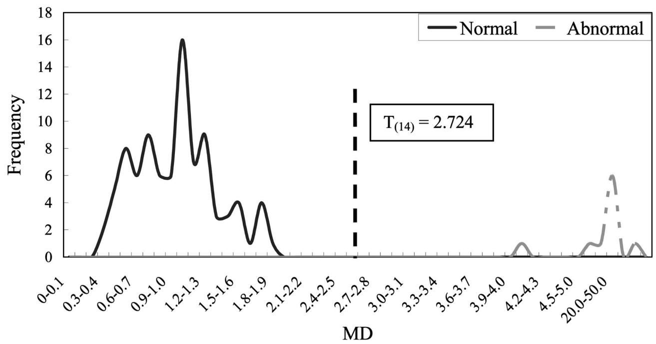

Fig. 5. MD distributions of the RF functional inspection test set (reduced model with 14 attributes).

TABLE 11
Results of All Reduced Models

Gain	# Attributes	Treshold	Accuracy	
			Training	Test
>0	36	1.994	99.67%	99.00%
>0.1	31	2.0345	99.00%	99.00%
>0.2	27	2.099	100.00%	100.00%
>0.3	24	2.1388	100.00%	100.00%
>0.4	17	2.5479	100.00%	100.00%
>0.5	**14**	**2.724**	**100.00%**	**100.00%**
>0.6	10	2.745	99.33%	99.00%
>0.7	8	3.4866	99.33%	98.00%
>0.8	7	1.791	87.00%	90.00%
>0.9	6	1.9867	87.67%	89.00%

5.4 Classification Robustness Evaluation

To demonstrate the robustness of MTS and the effectiveness of the proposed PTM for real applications, we also designed three experiments on this case data and compared the performance of MTS with that of SDA, C4.5, BPN, and SVMs. The data used for the analysis are described in Table 12. The training and test sets of experiment 1 were the same as that used in Section 5.2. From experiment 1 to experiment 3, the degree of class imbalance gradually increased.

The results are shown in Table 13. It is worthy to note that the specificities of SDA, C4.5, BPN, and SVMs were almost perfect in the three experiments, whereas their sensitivities performed poorly. Besides, the performance of our proposed PTM was better than that of the ESM in implementing MTS. MTS had the best classification ability and was the most robust classification technique.

6 Conclusion

MTS is a collection of methods proposed for a diagnostic and forecasting technique. MTS can be used not only to execute classification tasks but also to identify the important variables of a multivariate system. For a multivariate classification problem, the ratio of the positive to the negative examples of the training set is one of the important factors that impact on the effective training of a classifier. One of the main purposes of this study was to investigate the effect caused by an imbalanced training set on MTS and other popular classification techniques such as SDA, C4.5, BPN, and SVMs. Additionally, when implementing MTS in real applications, what seems to be lacking is a method to set a threshold efficiently to discriminate between the two classes and to avoid overfitting problems. Thus, a PTM was developed using Chebyshev's theorem. This study compared the performance of MTS with other classification techniques and also verified the performance of the PTM. The results indicated that the class imbalance problem indeed caused a training bias on SDA, C4.5, BPN, and SVMs, and it follows that the learned classifiers tended to weaken the ability to predict the minority class. In contrast, MTS performed robustly against the class imbalance

TABLE 12
RF Functional Inspection Data Set

Data Set	Class Maj. / Min.	# Var.	Exp.	*f*-ratio	Training Set		Test Set	
					# Neg.	# Pos.	# Neg.	# Pos.
RF Functional Inspection	Normal / abnormal	62	1	0.40	270	30	90	10
			2	0.38	450	30		
			3	0.39	810	30		

TABLE 13
Comparison of Test Results on the RF Functional Inspection Data Set

Data Set	Exp.	Index	MTS(PTM)	MTS(ESM)	SDA	C4.5	BPN	SVMs
RF	1	Sensitivity	100.00	100.00	40.00	60.00	60.00	80.00
		Specificity	100.00	100.00	100.00	100.00	100.00	100.00
		g-means	100.00	100.00	63.25	77.46	77.46	89.44
		RS	1.00	1.00	0.40	0.60	0.60	0.8
	2	Sensitivity	100.00	90.00	40.00	30.00	60.00	70
		Specificity	98.89	100.00	96.67	96.67	100.00	100
		g-means	99.44	94.87	62.18	53.85	77.46	83.67
		RS	1.01	0.90	0.41	0.31	0.60	0.7
	3	Sensitivity	90.00	80.00	40.00	30.00	50.00	80.00
		Specificity	100.00	100.00	100.00	100.00	100.00	100.00
		g-means	94.87	89.44	63.25	54.77	70.71	89.44
		RS	0.90	0.80	0.40	0.30	0.50	0.8

problem regardless of the data sizes and complexity of the problems. Besides, by using the PTM, the threshold was determined more efficiently, and the classification ability of MTS was enhanced. Finally, this study sought the assistance of the robustness of MTS to improve the RF functional inspection process of mobile phone manufacturing. The case study aimed to remove the redundant inspection attributes and expected to maintain the original inspection accuracy level. The results showed that the number of attributes was successfully reduced from 62 to 14 without losing inspection effectiveness. As a result of diminishing the inspection time and relative production expenses, the manufacturer improved its productivity and enhanced the flexibility of responding to demands; thus, the manufacturer stands on a vantage point vis-à-vis its competitive market. These results led us to conclude that MTS is a powerful and useful classification technique for imbalanced data analysis. Thus, it adequately appears that MTS is practical for industry improvement, especially for high-technology industries that have a high yield rate and with data that are imbalanced.

This study has four contributions. First, the research provided a detailed description of the concepts, principles, and implementing procedures of MTS. This can help us understand this diagnostic and forecasting technique. Second, with the systematic experiments and analyses, this paper showed that the imbalanced training set indeed causes a bias on classifier training, whereas MTS is robust even though the training set distribution is much imbalanced. This finding can be a reference for future classification tasks on imbalanced data. Third, this study proposed a PTM that can aid MTS in efficiently determining a threshold. Finally, we successfully introduced MTS to improve the RF functional inspection process of mobile phone manufacturing. This indicates the practicability of MTS in high-technology production.

However, there are some important issues that need to be studied in our future work. In a multivariate system, the case in which all samples have equal value on some variable and the multicollinearity problem are unavoidable. The former will result in a zero standard deviation on the variable, whereas the latter will cause some errors when establishing the correlation matrix. All these phenomena will lead to the unsuccessful construction of MS. Moreover, how to extend the MTS and the research scheme to multiclass tasks is also one of the important research topics that will be tackled in the near future.

REFERENCES

[1] N. Japkowicz, "Learning from Imbalanced Data Sets: A Comparison of Various Strategies," *Proc. Assoc. Advancement of Artificial Intelligence Workshop Learning from Imbalanced Data Sets (AAAI '00)* pp. 10-15, 2000.

[2] N. Japkowicz and S. Stephen, "The Class Imbalance Problem: A Systematic Study," *Intelligent Data Analysis,* vol. 6, no. 5, pp. 429-450, 2002.

[3] C. Phua, D. Alahakoon, and V. Lee, "Minority Report in Fraud Detection: Classification of Skewed Data," *SIGKDD Explorations,* vol. 6, no. 1, pp. 50-59, 2004.

[4] N.V. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," *J. Artificial Intelligence Research,* vol. 16, pp. 231-357, 2002.

[5] J.W. Grzymala-Busse, J. Stefanowski, and S. Wilk, "A Comparison of Two Approaches to Data Mining from Imbalanced Data," *Lecture Notes in Computer Science,* vol. 3213, pp. 757-763, 2004.

[6] P.C. Pendharkar, J.A. Rodger, G.J. Yaverbaum, N. Herman, and M. Benner, "Association, Statistical, Mathematical and Neural Approaches for Mining Breast Cancer Patterns," *Expert System with Applications,* vol. 17, pp. 223-232, 1993.

[7] M.A. Maloof, "Learning When Data Sets Are Imbalanced and When Costs Are Unequal and Unknown," *Proc. 20th Int'l Conf. Machine Learning Workshop Learning from Imbalanced Data Sets II (ICML '03),* 2003.

[8] G. Batista, R.C. Prati, and M.C. Monard, "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data," *SIGKDD Explorations,* vol. 6, no. 1, pp. 20-29, 2004.

[9] H. Guo and H.L. Viktor, "Learning from Imbalanced Data Sets with Boosting and Data Generation: The Databoost-IM Approach," *SIGKDD Explorations,* vol. 6, no. 1, pp. 30-39, 2004.

[10] N.V. Chawla, N. Japkowicz, and A. Kolcz, "Editorial: Special Issue on Learning from Imbalanced Data Sets," *SIGKDD Explorations,* vol. 6, no. 1, pp. 1-6, 2004.

[11] K. Huang, H. Yang, I. King, and M. Lyu, "Learning Classifiers from Imbalanced Data Based on Biased Minimax Probability Machine," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '04),* pp. 558-563, 2004.

[12] J. Zhang and I. Mani, "kNN Approach to Unbalanced Data Distributions: A Case Study Involving Information Extraction," *Proc. 20th Int'l Conf. Machine Learning Workshop Learning from Imbalanced Data Sets (ICML '03),* 2003.

[13] G. Taguchi, S. Chowdhury, and Y. Wu, *The Mahalanobis-Taguchi System.* McGraw-Hill, 2001.

[14] G. Taguchi and R. Jugulum, *The Mahalanobis-Taguchi Strategy.* John Wiley & Sons, 2002.

[15] W.H. Woodall, R. Koudelik, K.L. Tsui, S.B. Kim, Z.G. Stoumbos, and C.P. Carvounis, "A Review and Analysis of the Mahalanobis-Taguchi System," *Technometrics,* vol. 45, no. 1, pp. 1-15, 2003.

[16] A. Bovas and V. Asokan Mulayath, "Discussion—A Review and Analysis of the Mahalanobis-Taguchi System," *Technometrics,* vol. 45, no. 1, pp. 22-25, 2003.

[17] D.M. Hawkins, "Discussion—A Review and Analysis of the Mahalanobis-Taguchi System," *Technometrics,* vol. 45, no. 1, pp. 25-29, 2003.

[18] J. Rajesh, G. Taguchi, and S. Taguchi, "Discussion—A Review and Analysis of the Mahalanobis-Taguchi System," *Technometrics,* vol. 45, no. 1, pp. 16-21, 2003.

[19] J. Srinivasaraghavan and V. Allada, "Application of Mahalanobis Distance as a Lean Assessment Metric," *Int'l J. Advanced Manufacturing Technology,* vol. 29, pp. 1159-1168, 2006.

[20] T. Riho, A. Suzuki, J. Oro, K. Ohmi, and H. Tanaka, "The Yield Enhancement Methodology for Invisible Defects Using the MTS+ Method," *IEEE Trans. Semiconductor Manufacturing,* vol. 18, no. 4, pp. 561-568, 2005.

[21] P. Das and S. Datta, "Exploring the Effects of Chemical Composition in Hot Rolled Steel Product Using Mahalanobis Distance Scale under Mahalanobis-Taguchi System," *Computational Materials Science,* 2006.

[22] M. Kubat and S. Matwin, "Addressing the Curse of Imbalanced Training Set: One-Sided Selection," *Proc. 14th Int'l Conf. Machine Learning (ICML '97),* 1997.

[23] G. Wu and E.Y. Chang, "KBA: Kernel Boundary Alignment Considering Imbalanced Data Distribution," *IEEE Trans. Knowledge and Data Eng.,* vol. 17, no. 6, pp. 786-794, June 2005.

[24] R. Akbani, S. Kwek, and N. Japkowicz, "Applying Support Vector Machines to Imbalanced Datasets," *Proc. 15th European Conf. Machine Learning (ECML '04),* pp. 39-50, 2004.

[25] T.K. Ho and M. Basu, "Complexity Measures of Supervised Classification Problems," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 3, pp. 289-300, Mar. 2002.

[26] T. Evgeniou, M. Pontil, C. Papageorgiou, and T. Poggio, "Image Representations and Feature Selection for Multimedia Database Search," *IEEE Trans. Knowledge and Data Eng.,* vol. 15, no. 4, pp. 911-920, July/Aug. 2003.

[27] A. Sun, E.P. Lim, W.K. Ng, and J. Srivastava, "Blocking Reduction Strategies in Hierarchical Text Classification," *IEEE Trans. Knowledge and Data Eng.,* vol. 16, no. 10, pp. 1305-1308, Oct. 2004.

[28] G. Wu and E. Chang, "Class-Boundary Alignment for Imbalanced Dataset Learning," *Proc. 20th Int'l Conf. Machine Learning Workshop Learning from Imbalanced Data Sets II (ICML '03),* 2003.

[29] G. Wu and E. Chang, "Adaptive Feature-Space Conformal Transformation for Imbalanced Data Learning," *Proc. 20th Int'l Conf. Machine Learning (ICML '03),* pp. 816-823, 2003.

[30] C.T. Su, L.S. Chen, and T.L. Chiang, "A Neural Network Based Information Granulation Approach to Shorten the Cellular Phone Test Process," *Computers in Industry,* vol. 57, no. 5, pp. 379-390, 2006.

Chao-Ton Su received the PhD degree in industrial engineering from the University of Missouri, Columbia. He is a professor in the Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Taiwan, R.O.C. From 1985 to 2004, he was a faculty member in the Department of Industrial Engineering and Management, National Chiao Tung University, Taiwan. Since August 2006, he has been a Tsing Hua distinguished professor. His research activities include quality engineering and management, operation management, and neural networks in industrial applications. He is a member of the Institute of Industrial Engineers (IIE), Chinese Institute of Industrial Engineers (CIIE), and Chinese Society for Quality (CSQ) and a senior member of the American Society for Quality (ASQ). He is a two-time recipient (2000-2001 and 2002-2005) of the Outstanding Research Award from the National Science Council, Taiwan. In 2001, he also obtained the Individual Award of the National Quality Awards of the Republic of China.

Yu-Hsiang Hsiao received the BS and MS degrees in industrial engineering and management from the National Chiao Tung University, Taiwan, in 2003 and 2005, respectively. He is currently a PhD candidate in the Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Taiwan. His research interests include data mining, Six Sigma, and quality engineering.