International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.11, (2016), pp.147-160
http://dx.doi.org/10.14257/ijsip.2016.9.11.14

Accuracy Estimation of a Classifier Based on the Differences in Samples

Min Zhang[1] and Shengbo Yu[2]

[1]Software Theory and Technology Chongqing Key Lab
College of Computer Science
Chongqing University
[2]Software Theory and Technology Chongqing Key Lab
College of Computer Science
Chongqing University
[1]zm@cqu.edu.cn, [2]ysb@cqu.edu.cn

Abstract

The classification accuracy is an important standard to measure the quality of the classifier. Usually, the classification accuracy is assessed later, not during the classification process. Problems such as classification accuracy drops cannot be timely and effectively found. It is necessary that marking test samples when estimating classification accuracy. It is a problem that we care about that how much is the classification accuracy when a group of new samples obtained. The problem must be concerned when using and improving the classifier in the case of growing data. To solve this problem, this paper put forward different estimates from different perspectives which based on the difference between samples. One estimate is based on the difference in samples distribution, which is from the Bayesian criterion. Another estimate is based on the difference in each sample instance, which is from the K nearest neighbor classification. Classification accuracy is also estimated by using the artificial neural networks, which combine the characteristics of the above two methods. And results show the proposed methods have good effects.

Keywords: *Classification accuracy, Samples distribution, Bayesian criterion, K nearest neighbor classification, Artificial neural networks*

1. Introduction

The accuracy estimation of classifiers [1] is an important issue in machine learning. At present, classification accuracy has been estimated by holdout, random sub-sampling, k-fold cross-validation [2] and so on. A new classification accuracy estimation method combined the advantages of random sub-sampling and k-fold cross-validation was put forward in [3]. A bootstrap method for assessing classification accuracy and confidence for agriculture land use mapping was introduced in [4-5] investigated using the meta-learning for predicting the accuracy of a classifier in text classification task. The method involved extracting a descriptive feature set for each token of the free text data and then capturing their distribution via the meta-features.

At present, the classification accuracy has been obtained by two common methods. One common method is observing the training accuracy. The high training accuracy is not the sufficient and necessary condition of high classification accuracy. However, estimating the test accuracy by observing the training accuracy also has some significance. The test accuracy is often lower than the training accuracy. If the training accuracy of a classifier is too low, then the test accuracy of the classifier is also not high. Another common method is marking the test samples. We can get the test accuracy of a group of labeled test samples by comparing the classification result. And using the statistics knowledge, the test accuracy

ISSN: 2005-4254 IJSIP

will be extended to the general case. This method requires marking some samples. Sometimes, labeling samples will be a little expensive. However, the obtained estimated value of classification accuracy estimation is object. So, if we want to get the classification accuracy, we often need to label the samples. Can we not labeling additional samples to get the classification accuracy?

This paper proposes the classification accuracy estimation methods which base on the difference between training samples and test samples. The differences between the samples can be considered from the perspective of macro and micro. The differences of samples distribution are the differences of samples from macro perspective. And the differences of samples instances are the differences of samples from micro perspective. We can get the classification accuracy estimation by comparing the test samples and training samples. The methods can avoid the expensive cost of labeling samples. It can also make the machine learning system giving classification accuracy as well as classification results. The methods can be applied to online learning, transfer learning and so on.

2. Classification Accuracy Estimation Based on the Differences in Samples Distribution

For results of the classifiers, according to the Bayesian criterion [6], there is

$$\mathrm{p}(\mathrm{y}|\mathrm{x}) = {}^{P(xy)}/_{P(x)} \tag{1}$$

where posterior probability $\mathrm{P}(\mathrm{y}|\mathrm{x})$ represents that label y is the label of sample x, while x be known. The probability $\mathrm{P}(\mathrm{xy})$ represents that label of sample x is y. The prior probability $\mathrm{P}(\mathrm{x})$ represents the probability of sample x. According to the Equation (1), it can be seen that if $\mathrm{P}(\mathrm{xy})$ is a constant, when $\mathrm{P}(\mathrm{x})$ becomes bigger, $\mathrm{P}(\mathrm{y}|\mathrm{x})$ becomes smaller. And when $\mathrm{P}(\mathrm{x})$ becomes smaller, $\mathrm{P}(\mathrm{y}|\mathrm{x})$ becomes bigger. The classification results of the Bayesian classifier have been direct impacted by the changes of $\mathrm{P}(\mathrm{x})$. In general, classifiers assume that $\mathrm{P}(\mathrm{x})$ does not change during the classification process. It is often means that samples distribution changes if $\mathrm{P}(\mathrm{x})$ changes. There are some differences of distribution between the old and new samples. The classifier, which is obtained by the old training samples, is not applicable to the new test samples. Otherwise, it will lead to decline classification accuracy. So how do we measure the differences between two samples distribution?

2.1. MMD Statistics

A set of training samples is marked as $\mathrm{A}(x_1, x_2, \dots, x_m)$, which obey the distribution p. And a set of test samples is marked as $\mathrm{B}(y_1, y_2, \dots, y_n)$, which obey the distribution q. How to determine p and q are the same. In the past, it is mainly used parametric statistical methods. Firstly, it is need to determine their distributions model. And then, infer whether they contain the same parameters by using the method of parameter hypothesis. [7] presented a method that a Hilbert space embedding for distribution in 2007. [8] proposed a kernel method for measuring samples distribution difference between two groups samples in 2007, which be called Maximum Mean Discrepancy (MMD). The MMD statistics is defined as follows:

$$\mathrm{MMD}[F, p, q] = \left|\frac{1}{m^2}\sum_{i,j=1}^{m} k(x_i, x_j) - \frac{2}{mn}\sum_{i,j=1}^{i=m,j=n} k(x_i, y_j) + \frac{1}{n^2}\sum_{i,j=1}^{n} k(y_i, y_j)\right|^2 \tag{2}$$

where $\mathrm{k}(,)$ is the kernel function. And the $\mathrm{MMD}^{\mathrm{T}}$ statistics is defined as follows:

$$\mathrm{MMD}^{\mathrm{T}}[F, p, q] := \frac{\mathrm{MMD}[F,p,q]}{\sqrt{2K}} \tag{3}$$

where F is a function which mapping the measuring volume to the real domain and K is a constant. In $\mathrm{MMD}^{\mathrm{T}}$, there have $|\mathrm{k}(\mathrm{x},\mathrm{y})| \le \mathrm{K}, \mathrm{x} \in \mathrm{A}, \mathrm{y} \in \mathrm{B}$.

The computational complexity of the MMD statistics is $\mathrm{O}(m+n)^2$. According to the sampling statistical knowledge, under the α level of hypothesis testing, the acceptance region of $\mathrm{MMD}^{\mathrm{T}}[F,p,q]$ is $[0,1+\sqrt{2loga^{-1}})$.

The MMD can calculate the differences of two groups of samples distribution. While condition of the MMD is that A and B are independent random variables. The greater the value of MMD, the greater the differences between test samples distribution and training samples distribution. And then the classification accuracy of the test samples is lower. The smaller the value of MMD, the smaller the difference between test samples distribution and training samples distribution. And then the classification accuracy of the test samples is higher. The MMD is an unbiased U statistics. It is limited by samples size. When the samples size is small, representative information of each sample is accordingly increased. And the influence of each sample difference to samples distribution difference will be magnified.

2.2. MMD Statistics and Classification Accuracy

In order to determine the relationship between the MMD and the classification accuracy, we use the parameter estimation method to estimate classification accuracy. From the experimental data of the MMD and classification accuracy, it can be seen that the MMD and classification accuracy have good linearity. The problem of classification accuracy estimation can be assumed the equation $\mathrm{TAE(MMD)} = \mathrm{f1(MMD, C1)}$. The TAE, test accuracy estimate, is the classification accuracy estimation of test samples. The symbol $\mathrm{C1}$ is the list of parameters. The symbol $\mathrm{f1}$ is the linear function. In order to determine the parameters C1 and minimize the error of the test accuracy and test accuracy estimation, we use the least square method [9]. There are z1 groups of experimental data as follows: $(\mathrm{mmd_i}, \mathrm{y_i}), \mathrm{i} = 1, \ldots, \mathrm{z}1$. And they are independent of each other. The $\mathrm{mmd_i}$ and $\mathrm{y_i}$, respectively, are the value of MMD and test accuracy of the experiment i-th. We have the $\mathrm{X} = \left| \vdots \quad \cdots \right|$, $\mathrm{C1} = [\;]$, $\mathrm{Y} = [\cdots]$.

The sum of squared residuals of y-axis is given as follows:

$$\mathrm{Q_2} = ||\mathrm{Y} - \mathrm{XC1}||_2 = (\mathrm{Y} - \mathrm{XC1})'(\mathrm{Y} - \mathrm{XC1}). \tag{4}$$

According to the vector differential theory, taking the derivative with respect to C1 on $\mathrm{Q_2}$ and have the equation $\frac{\partial \mathrm{Q_2}}{\partial \mathrm{C1}} = -2\mathrm{X}'(\mathrm{Y} - \mathrm{XC1})$. For practical significance, the optimal solution of the problem is given as follows: $\widetilde{\mathrm{C1}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Y}$.

3. Classification Accuracy Estimation Based on the Differences in Samples Instances

The MMD statistics is limited by samples size. When dealing with a small number of samples, it is not reliable. Because the influence that the difference of each sample to the difference of samples distribution will be magnified. Therefore, we propose the classification accuracy estimation method based on the differences in sample instances.

3.1. MMR Statistics

K-Nearest-Neighbor classifier [10] (kNN) is based on the sample instances. It means that each sample can be represented by its k nearest neighbors. Its basic idea is that considering the labels of k nearest neighbors of a sample instance, if most of the labels are a category, then the label of the sample instance is the category. Let us see the mathematical model of the kNN. A training samples set is marker as $\mathrm{A}(x_1, x_2, \ldots\ldots, x_m)$, and a test samples set is marker as $\mathrm{B}(y_1, y_2, \ldots\ldots y_n)$. In order to obtain the label of each sample instance $\mathrm{y_i}$ in B, it is necessary to get the k nearest neighbors of $\mathrm{y_i}$. And then getting the label of sample $\mathrm{y_i}$ is by voting from the k nearest neighbors. An obvious improvement of the kNN is to distance weighted on k nearest neighbors [11]. For k nearest neighbors, the more close to the test sample, the greater the weight. It can be seen that if the distance between the nearest neighbors and the test samples is smaller, the higher the credibility of the classification result. The accuracy of the kNN classification results can be estimated by the distance between the nearest neighbors and the test samples. The smaller the distance, the more reliable classification result. Here, we choose the special case $k = 1$. For each sample y_i in test samples B, if the distance between y_i and its nearest sample x_i in training samples A is smaller, the higher accuracy y_i and x_i have the same label. On the contrary, if the distance between y_i and x_i is bigger, the lower accuracy y_i and x_i have the same label.

Assumed that the minimum distance between samples in A and B can affect accuracy of the kNN classification results. By observing the minimum distance between each sample in B and samples in A, the classification accuracy estimation about a set of samples in B can be obtained. Through the accuracy estimation of the kNN, the classification accuracy estimation of general classifier can be acquired. This paper proposes the Mean Maximum Resemblance (MMR) statistics, which is the mean minimum distance between each sample in B and samples in A. So we have the equation

$$\mathrm{MMR}(A, B) := \mathrm{mean}(\min_{x \in A} d(x, y)). \tag{5}$$

The MMR calculation process is as follows:

1. For $y_i \in B, i = 1,2,\dots,n$, computing the minimum distance between y_i and the training samples A by the following formula.

$$md(y_i) = \min_{x_i \in A} d(x_i, y_i) = \min_{x_i \in A} (x_i - y_i) * (x_i - y_i)', j = 1,2,\dots,m. \tag{6}$$

2. Compute the mean of $md(y_i)$ by the following equation.

$$md(B) = \frac{\sum_{i=1} md(y_i)}{}. \tag{7}$$

3. Standardize the $md(B)$ using the following formula.

$$stdd(B) = \frac{md(B)}{(d(A, A))}$$

$$\max(d(A, A)) = \max_{x_{j_1} \in A, x_{j_2} \in A} \left(d(x_{j1}, x_{j2})\right) = \max_{x_{j_1} \in A, x_{j_2} \in A} (x_{j1} - x_{j2}) * ((x_{j1} - x_{j2}))', j_1, j_2 = 1,2,\dots,m \tag{8}$$

In step 3, we use the maximum span as normalization denominator. Its goal is that making the MMR statistics is not affected by samples as soon as possible. The greater the value of MMR, the greater the difference between the test samples instances and the training samples instances. And then the classification accuracy of test samples is lower. The smaller the value of MMR, the smaller the difference between the test samples instances and the training samples instances. And then the classification accuracy of the test samples is higher.

The MMR statistics has some properties. The equation $\mathrm{MMR}(A, B) = 0$ is true if and only if samples set A contains samples set B. For each sample instance in test samples, it can be found in the training set B, which can make the equation $\mathrm{MMR}(A, B) = 0$ established. The $\mathrm{MMR}(A, B)$ statistics is not symmetrical. In other words, the equation $\mathrm{MMR}(A, B) \neq \mathrm{MMR}(B, A)$ is true. For example, when the set A contained within the set B, the equations $\mathrm{MMR}(A, B) = 0$ and $\mathrm{MMR}(B, A) \neq 0$ are true. The computation complexity of the MMR is $O(mn + m^2)$.

3.2. MMR Statistics and Classification Accuracy

In order to determine the relationship between the MMR and the classification accuracy, we use the parameter estimation method to estimate classification accuracy. From the experimental data of the MMR and the classification accuracy, it can be seen that the MMR and classification accuracy have good linearity. Therefore, liking the problem that the classification accuracy estimation based on the differences in samples distribution, the problem of classification accuracy estimation based on the differences in sample instances can be assumed that the equation $\mathrm{TAE(MMR)} = \mathrm{f2(MMR, C2)}$. The TAE(MMR) is the classification accuracy estimation of test samples. The symbol $\mathrm{C2}$ is the list of parameters. The symbol $\mathrm{f2}$ is the linear function. In order to determine the parameters C2 and minimize the error of the test accuracy and test accuracy estimation, we use the least square method also.

4. Classification Accuracy Estimation Based on the MMD and MMR

The MMD focuses on the differences between training samples distribution and test samples distribution, which is a statistics from macro statistics. The MMR focus on the differences between training samples instances and test samples instances, which is a statistics from micro statistics. They can complement with each other, and jointly estimate the classification accuracy of test samples. Therefore, we use the artificial neural networks [12-13], which combine the MMD and MMR, to estimate the test accuracy.

The artificial neural networks are suitable the problems, such as the training samples containing noise, samples instances having a lot of attributes and so on. They are also suitable the problem that represented by a lot of symbolic. The artificial neural networks are also a kind of function approximation method, which can simulate the complex function relation. In other words, they can effectively simulate the complex relation between the input and output of samples. They have self-organization ability. In the training process of the artificial neural networks, the input data after handled by the artificial neural networks can produce the output data. By adjusting the parameters of each node in the artificial neural networks to minimizes the error between the output and the actual output data of the input.

Extreme learning machine (ELM) [14-15] is a kind of fast single hidden layer artificial neural networks algorithm. The characteristics of the algorithm are that parameters of the hidden layer nodes, named the weight and bias value obtained randomly without regulated in the process of determining neural networks parameters. And the weights of the output layer obtained by minimizing the squared loss function. The process of determining parameters of the neural networks is without any iteration steps, which can greatly reduces the time of adjusting the networks parameters.

Formally, the N independent data samples are given by $\{(x_i, t_i)\}_{i=1}^{N} \subset \mathbb{R}^2 \times \mathbb{R}^1$ with $x_i = \begin{bmatrix} \mathrm{mmd}_i \\ \mathrm{mmr}_i \end{bmatrix}$. The parameters mmd_i and mmr_i are the MMD and MMR of i-th data sample, respectively. The symbol t_i is the classification accuracy of i-th data sample. The hidden layer of the ELM neural networks has L neurons. The parameters a_i and b_i are the weight and bias value of i-th hidden layer node, respectively. The symbol β_i is the weight of the connection between the i-th hidden layer node and the output layer. $G(a_i, b_i, x)$ is the output of the i-th hidden layer node corresponding to the sample x. The square loss function is as follows:

$$J = (H\beta - T)'(H\beta - T) \tag{9}$$

where $H = \begin{bmatrix} & \vdots & \ddots & \vdots & \end{bmatrix}_{N\times L}$, $\beta = \begin{bmatrix} \vdots \end{bmatrix}_{L\times 1}$ and $T = \begin{bmatrix} \vdots \end{bmatrix}_{N\times 1}$. The problem of determining the parameters of neural networks is converted into minimizing the square loss function. That is to say, finding the parameter $\ddot{\beta}$ making the equation $||H\ddot{\beta} - T|| = \min_{\beta} ||H\beta - T||$ established.

5. Experiments and Analysis

5.1. Experimental Settings

There are three data sets in the experiment, named experimental Data 1, experimental Data 2 and experimental Data 3, respectively.

The experimental Data 1 is the academy of agricultural sciences' ulcer recognition system data. It has two groups of data samples, collected in 2010 and 2012, respectively.

The first data samples, collected in 2010, have 891 samples instances. The second data samples, collected in 2012, have 90 samples instances. The ulcer lesions have some changes in the two collections.

The experimental Data 2 is the ICDM 2007 data mining competition task 2 data, which is related to the problem of WiFi indoor positioning. This data is used to transfer learning. It has two groups of data samples, collected in two different time periods, respectively. The distribution of the two groups of data samples is totally different, which is the extreme case of samples change. The Task_2_landmark_ -data set, collected in the easier time periods, is the data of source task in transfer learning. The Task_2_test_data set, collected in the later time periods, is the data of target task in transfer learning.

The experimental Data 3, which from UCI machine learning repository, is the Wilt data set. It is about the tree withered. The training set has 4339 samples instances. And the test set has 500 samples instances.

This paper carries out 40 times experiment, named Task1 to Task40, respectively. The classification accuracy of the training samples and test samples is obtained by the ELM classifier in this paper.

5.2. Experiment One: Samples Difference, MMD and MMR

The Task1, Task2 and Task3 use the same training samples to get the same classifier. The training samples set is randomly selected 801 samples instances from the first data samples in the experimental Data 1. The test samples set of the Task1 is the total samples instances of the first data samples expect the training samples in the experimental Data1. The test samples set of the Task2 is the total samples instances of the second data samples in the experimental Data1. The test samples of the Task3 are randomly selected half of the test samples of the Task1 and half of the test samples of the Task2, which simulate the intermediate statuses between the first data samples statuses and the second data samples statuses. Table 1 shows the results of the Task1 to Task3. The accuracy statistics is test accuracy divided by training accuracy.

Table 1. Results of the Task1 to Task3

Experimental number	Training accuracy	Test accuracy	Accuracy statistics	MMD	MMR(*0.01)
Task 1	0.9388	0.9444	1.0060	0.2668	0.1809
Task 2	0.9388	0.6556	0.6983	4.1423	4.0663
Task 3	0.9388	0.7667	0.8167	1.9410	2.0966

The Table 1, shows that the first data samples and the second data samples are different. And the expression of ulcer lesions is different in the two collections.

The data set of the Task4 to Task10 is the experimental Data 2. The Task4 to Task10 use the same training samples set, which is randomly selected 1200 samples instances from the Task_2_landmark_data set. The test samples of the Task4 are randomly selected 800 samples instances from the Task_2_lamdmark_data set expect the training samples. The test samples of the Task5, Task6 and Task7 are randomly selected 800 samples instances from the Task_2_test_data set, respectively. The test samples of the Task8, Task9 and Task10 are randomly selected half of the test samples of the Task5, Task6 and Task7 and half of the test samples of the Task4, respectively. Table 2, shows the results of the Task4 to Task10.

Table 2. Results of the Task4 to Task10

Experimental number	Training accuracy	Test accuracy	Accuracy statistics	MMD	MMR(*0.01)
Task 4	0.8989	0.8900	0.9901	1.0566	1.0236
Task 5	0.8989	0.0025	0.0028	6.7068	25.3677
Task 6	0.8989	0.0050	0.0056	6.6975	25.4288
Task 7	0.8989	0.0038	0.0042	6.6564	25.3137
Task 8	0.8989	0.4475	0.4978	3.5084	13.1746
Task 9	0.8989	0.4325	0.4811	3.3960	13.2386
Task 10	0.8989	0.4600	0.5117	3.4973	13.1392

The Table 2, shows that in the problem of WiFi indoor positioning, the collected samples in different time periods have different expression.

The data set of the Task11 to Task17 is the experimental Data 3. The Task11 to Task17 use the same training samples, which are randomly selected 1200 samples instances from the training set in the experimental Data 3. The test samples set of Task11 is randomly selected 300 samples instances from the training set expect the training samples. The test samples of Task12 to Task14 are randomly selected 300 samples instances from the test set in the experimental Data 3, respectively. The test samples of the Task15, Task16 and Task17 are randomly selected half of the test samples of the Task12, Task13 and Task14 and half of the test samples of the Task11, respectively. Table 3, shows the results of the Task11 to Task17.

Table 3. Results of the Task11 to Task17

Experimental number	Training accuracy	Test accuracy	Accuracy statistics	MMD	MMR(*0.01)
Task 11	0.9825	0.9800	0.9975	0.4331	0.0606
Task12	0.9825	0.6233	0.6344	6.2932	0.9690
Task13	0.9825	0.6267	0.6378	6.2790	0.9921
Task14	0.9825	0.6033	0.6141	6.2235	1.0134
Task15	0.9825	0.8033	0.8176	3.3556	0.5166
Task16	0.9825	0.8033	0.8176	3.2629	0.5233
Task17	0.9825	0.7767	0.7905	3.2629	0.5821

The Table 3, shows that the training set and test set in the Wilt data set are different. The samples, collected in the same period, follow the same distribution. The old samples follow the old distribution. And the new samples follow the new distribution. The samples, mixed by the old and new samples, can be seen as transitional phase data set.

The experimental results show that the MMD can correctly reflect the difference of samples distributions. The smaller the value of MMD shows that test accuracy is closer to training accuracy. According to the above discussion of this paper, the smaller the value of MMD shows that the difference between training samples distribution and test samples distribution is smaller. And the smaller difference between training samples distribution and test samples distribution shows that test accuracy is closer to training accuracy. The experimental results and the previous inference of the MMD in this paper are same. The experimental results also show that the MMR can correctly reflect the difference between training samples instances and test samples instances. The greater the value of MMR shows that test accuracy is smaller. According to the previous discussion of this paper, the greater the value of MMR shows that the minimum distance between test samples instances and training samples instances is greater, which shows that the test accuracy is smaller. The experimental results and the previous inference of MMR in this paper are same.

5.3. Experiment Two: Samples Size, MMD and MMR

The data set of the Task18 to Task26 is the experimental Data 2. The Task18 to Task26 use the same training samples, which are randomly selected 1200 samples instances from the Task_2_landmark_data set. The test samples of the Task i $(i = 18,19, \ldots 26)$ are randomly selected $(10 * (i - 17) * (i - 17))$ samples instances from the Task_2_landmark_data set expect the training samples, respectively. Table 4, shows the results of the Task18 to Task26.

Table 4. Results of the Task18 to Task26

Experimental number	Test samples size	Accuracy statistics	MMD	MMR （*0.01）
Task 18	10	1.0126	1.4972	1.2251
Task 19	40	0.9282	1.3811	1.1204
Task 20	90	0.9126	1.3024	1.2813
Task 21	160	1.0056	1.2426	1.2872
Task22	250	1.0036	1.2262	1.2946
Task 23	360	1.0126	1.2080	1.3083
Task 24	490	0.9850	1.1880	1.2288
Task 25	640	1.0267	1.1764	1.2024
Task 26	810	1.0320	1.1681	1.2049

The data set of the Task27 to Task35 is the experimental Data 3. The Task27 to Task35 use the same training samples, which are randomly selected 1200 samples instances from the training set. The test samples of the Task i $(i = 27,28, \ldots 35)$ are randomly selected $(10 * (i - 26) * (i - 26))$ samples instances from the training set expect the training samples, respectively. Table 5, shows the results of the Task27 to Task35.

Table 5. Results of the Task27 to Task35

Experimental number	**Test samples size**	**Accuracy statistics**	**MMD**	**MMR （*0.01）**
Task 27	10	1.0178	1.5171	0.0199
Task 28	40	0.9924	1.4173	0.0201
Task 29	90	0.9944	1.3420	0.1982
Task 30	160	1.0178	1.2859	0.1999
Task31	250	1.0015	1.2438	0.0213
Task 32	360	1.0008	1.2168	0.0219
Task 33	490	1.0074	1.1935	0.0222
Task 34	640	0.9955	1.1809	0.0244
Task 35	810	0.9990	1.1702	0.0211

The Table 4, and Table 5, show that when the training samples size is enough, the MMD is limited by test samples size. The larger the test samples size, the more accurate the results. When the test samples size is small, the accuracy of the MMD results is low. The accuracy statistics fluctuates near 1 in the Table 5, which shows that test accuracy and training accuracy are almost equal. When the test samples size is enough, for example the Task25 and Task34, the MMD is less volatile. When the test samples size is small, such as the Task18 and Task27, the MMD is volatile. However, the accuracy statistics is still around 1. And then it is not reliable to use the MMD statistics to estimate the test accuracy. And these experimental results also show that samples size has less effect on the MMR, which is based on the samples instances.

5.4. Experiment Three: Test Accuracy Estimation

The Task36 to Task40 use the MMD statistics, the MMR statistics and the artificial neural networks, which combine with the MMD and MMR to estimate the test accuracy.

The data of the Task36 is the experimental Data 1. The training samples set of the Task36 is randomly selected 801 samples instances from the first data samples. The test samples set of the Task36 is randomly selected 90 samples instances from the experimental Data 1 expect the training samples. Repeat 20 times, and then get the test accuracy, MMD statistics and MMR statistics. The 15 group data in front of the experimental data of the Task36 are training samples, and the other 5 group experimental data are test samples. They are used to estimate the test accuracy. Table 6, shows the results of test accuracy estimation experiment. The TAE(MMD) is the test accuracy estimation by using the MMD and linear estimation. The TAE(MMR) is the test accuracy estimation by using the MMR and linear estimation. The TAE(MMD,MMR) is the test accuracy estimation by using the ELM neural networks.

Table 6. Results of the Task36

Experimental number	Test accuracy	MMD	MMR(*0.01)	TAE(MMD)	TAE(MMR)	TAE(MMD,MMR)
Task36	0.8667	1.5719	1.4702	0.8264	0.8271	0.8531
	0.7111	2.8547	2.8307	0.7236	0.7135	0.6306
	0.9111	0.3444	0.4123	0.9248	0.9153	0.9121
	0.6778	3.9338	4.5207	0.6371	0.5725	0.6291
	0.7444	1.9269	1.6945	0.7979	0.8083	0.7892

Table 6, shows that the average error of test accuracy by using the MMD is 0.0321. And the average error of test accuracy by using the MMR is 0.0431. The average error of test accuracy byusing the ELM neural networks estimation is 0.0377.

The data set of the Task37 and Task38 is the experimental Data 2. The training samples of the Task37 and Task38 are randomly selected 801 and 1200 samples instances from the Task_2_landmark_data set, respectively. The test samples of the Task37 and Task38 are randomly selected 90 and 250 samples instances from the experimental data 2 expect the training samples, respectively. Repeat 20 times, and then get the test accuracy, MMD and MMR, respectively. The 15 group data in front of the experimental data of the Task37 and Task38 are training samples, respectively. And the other 5 group experimental data are test samples, respectively. They are used to estimate the test accuracy. Table 7, shows the results of test accuracy estimation experiment.

Table 7. Results of the Task37 and Task38

Experimental number	Test accuracy	MMD	MMR(*0.01)	TAE(MMD)	TAE(MMR)	TAE(MMD,MMR)
Task37	0.8667	1.3808	3.2139	0.7429	0.8424	0.8112
	0.1222	2.9120	22.8647	0.1146	0.1461	0.1103
	0.3778	2.3529	16.5390	0.3440	0.3702	0.3428
	0.0556	3.2491	24.8202	-0.0237	0.0768	0.0654
	0.0111	3.2577	27.1138	0.0792	-0.0045	0.0989
Task38	0.8440	1.3200	2.4968	0.7898	0.8495	0.8422
	0.1000	4.1463	23.5387	0.1278	0.0873	0.0980
	0.7640	1.5120	5.3105	0.7448	0.7476	0.7757
	0.3400	3.2696	16.4854	0.3331	0.3428	0.3755
	0.8160	1.2872	3.3255	0.7975	0.8195	0.8429

Table 7, shows that the average error of test accuracy by using the MMD in the Task37 and Task38 are 0.0625 and 0.0253, respectively. And the average error of test accuracy by using the MMR in the Task37 and Task38 are 0.0185 and 0.0082, respectively. The average error of test accuracy by using the ELM neural networks in the Task37 and Task38 are 0.0400 and 0.0156, respectively.

The data set of the Task39 and Task40 is the experimental Data 3. The training samples of the Task39 and Task40 are randomly selected 801 and 1200 samples instances from the training set, respectively. The test samples of the Task39 and Task40 are randomly selected 90 and 250 samples instances from the experimental Data 3 expect the training samples, respectively. Repeat 20 times, and then get the test accuracy, MMD statistics and MMR statistics, respectively. The 15 group data in front of the experimental data of the Task39 and Task40 are training samples, respectively. And the other 5 group experimental data are test samples, respectively. They are used to estimate the test accuracy. Table 8 shows the results of test accuracy estimation experiment.

Table 8. Results of the Task39 and Task40

Experimental number	Test accuracy	MMD	MMR(*0.01)	TAE(MMD)	TAE(MMR)	TAE(MMD,MMR)
Task39	0.6667	3.4588	1.1865	0.6365	0.7177	0.5968
	0.6222	3.4230	1.1510	0.6402	0.7243	0.5976
	0.8222	2.0074	0.7119	0.7830	0.8063	0.7818
	0.9222	1.1264	0.2927	0.8719	0.8845	0.8924
	0.7889	2.1589	0.5575	0.7677	0.8351	0.8035
Task40	0.7280	4.1775	0.7066	0.7307	0.7633	0.7278
	0.9520	0.4698	0.2307	0.9544	0.9164	0.9541
	0.8320	2.3866	0.3418	0.8387	0.8807	0.8576
	0.8560	1.8785	0.4962	0.8694	0.8310	0.8908
	0.7320	3.9761	0.7849	0.7428	0.7382	0.7195

Table 8, shows that the average error of test accuracy by using the MMD in the Task39 and Task40 are 0.0318 and 0.0072, respectively. And the average error of test accuracy by using the MMR in the Task39 and Task40 are 0.0506 and 0.0302, respectively. The average error of test accuracy by using the ELM neural networks in the Task39 and Task40 are 0.0359 and 0.0150, respectively.

According to the experimental results of the Task36, Task37 and Task39, it can be seen that for different data sets, the performance of the MMD and MMR is different. How to according to the problem to select the appropriate test accuracy estimation method and discussing the problem influences on classification accuracy estimation of the MMD and MMR are the research task in the next phase. Contrast the experimental results of the Task37, Task38 and Task39, Task40, it can be seen that the more the training samples and the test samples, the precision of the three test accuracy estimation methods is higher. For the test accuracy estimation, the average error by using the ELM neural networks is in the middle of the average error by using the MMD and MMR. These experimental results show that the test accuracy estimation by using the MMD, MMR and ELM neural networks are meaningful.

6. Conclusion

At present, there is no good estimation method for classification accuracy. In order to solve this problem, this paper put forward the classification accuracy estimation methods. The method of MMD statistics is based on the difference in sample distribution, which is from the Bayesian criterion. The method of MMR statistics is based on the difference in each sample instance, which is from the K nearest neighbor classification. The MMD statistics focuses on the difference between training samples distribution and test samples distribution, which is a statistics from macro perspective. However, it is limited by the samples size. The MMR statistics focuses on the difference between training samples instances and test samples instances, which is a statistics from micro perspective. It is associated with specific samples. And then classification accuracy is also estimated by using the ELM neural networks, which combines the characteristics of the MMD statistics and MMR statistics. The experimental results show that the three estimation methods have a good effect. They can be applied to online learning, transfer learning and so on. For different data sets, the performance of the MMD statistics and MMR statistics is different. So the research task in the next phase is how to select the appropriate classification accuracy estimation method according to the problem and to discuss the problems influence on classification accuracy estimation of the MMD and MMR.

Acknowledgments

This work is supported by the Fundamental Research Funds for the Central Universities (CDJZR20121805), National Natural Science Foundation of China (61502060) and Fundamental Research Funds for the Central Universities (CDJZR12180005).

References

[1] T. M. Mitchell, “Machine Learning “, McGraw-Hill Science, New York, **(1997)**

[2] P. N. Tan, “Introduction to Data mining”, Addison-Wesley, New Jersey, **(2006)**.

[3] Y. K. Wu, F. Duan and X. M. Yin, “Research on Accuracy Evaluation of Classifier”, Computer Development and Applications, vol. 24, no. 4, **(2010)**, pp. 10-15.

[4] C. Champagne, H. McNairn, B. Daneshfar and J. Shang, “A bootstrap method for assessing classification accuracy and confidence for agricultural land use mapping in Canada”, International Journal of Applied Earth Observation and Geoinformation, vol. 29, **(2014)**, pp. 44-52.

[5] S. Chaturvedi, T. A. Faruquie and L. V. Subramaniam, “Estimating Accuracy for Text Classification Tasks on Large Unlabeled Data”, Proceedings of the 19th International Conference on Information and Knowledge Management, Toronto, **(2010)** Octorber 26-30.

[6] H. Yu, Z. H. Zou and A. Y. Zou, “Machine Learning and Applications”, Tsinghua University Press, Beijing, **(2006)**.

[7] A. Smola, A. Gretton, L. Song and B. Scholkopf, “A hilbert space embedding for distributions”, Proceedings 18th International Conference, ALT, Sendai, **(2007)** Octorber 1-4.

[8] A. Gretton, M. K. Borgwardt, M. Rasch, B. Scholkopf and A. Smola, “A kernel method for the two-sample-problem”, Advances in Neural Information Processing Systems, Vancouver, **(2007)** December 4-7.

[9] S. Dong and T. Takayama, “improved Least Square Method for Selecting Design Wave Height”, Proceedings of the 12th International Offshore Polar Engineering Conference, vol. 12, **(2002)**, pp. 60-65.

[10] J. Han, M. Kamber and J. Pei, “Data Mining: Concepts and Techniques”, Morgan Kaumann Publishers, San Francisco, **(2006)**.

[11] N. Ishii, T. Murai, T. Yamada and Y. Bao, “Classification by weighting, similarity and kNN”, Lecture Notes in Computer Science, vol. 4224, **(2006)**, pp. 57-64.

[12] B. H. Demuth, M. Beale and M. T. Hagan, “Neural Network Design”, China Machine Press, Beijing, **(1995)**.

[13] S. Walczak and N. Cerpa, “Heuristic principles for the design of artificial neural networks”, Information and Software Technology, vol. 41, no.2, **(1999)**, pp. 107-117.

[14] G. B. Huang, Q. Y. Zhu and C. K. Siew, “Extreme learning machine: a new learning scheme of feed forward neural networks”, Neurocomputing, vol. 70, **(2006)**, pp. 489-501.

[15] G. B. Huang, D. H. Wang and Y. Lan, “Extreme learning machines: a survey”, International Journal of Machine Learning and Cybernetics, vol. 2, no. 2, **(2011)**, pp.107-122.

Authors

Min Zhang, she was born in Chongqing, China in 1978. She received the PhD degree in Computer Application from Chongqing University in 2008, M.S. degree in Software Theory from Chongqing University in 2003 and B.S. degree in Computing Sciences from Chongqing University in 2001. Now she works primarily in intelligent computation and learning machine. Her research interests include learning algorithm and pattern recognition.

Shengbu Yu, ea was born in Jianxi province, China 1990. He received his B.S. degree in Information Engineering from Nanchang Hangkong University, Jiangxi, China, 2013. Now he is pursuing his M.E. Degree in Software Theory and Technology Chongqing Key Lab, Computer College of Chongqing University, China. His research interests concentrate on pattern re