Investigation on enhancing the binary classification accuracy of supervised classifiers using various transforms

K.Issac1, K.Bharadwaj1, Bharanidharan.N1, Harikumar Rajaguru2
1,Department of ECE, Veltech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Chennai bharani2410@gmail.com
2Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam

Abstract. The Classification is used for testing instances where the unknown class labels are assigned where the predictor features are known. This paper aims to investigate the classification performance improvement of popular supervised classification approaches using data transformation techniques. Hilbert Transform, Discrete Wavelet Transform, and Principal Component Analysis are investigated as data transformation techniques for improving the performance of four different supervised classification approaches namely K-Nearest Neighbor classifier, Random Forest Classifier, Naive Bayes Classifier, and Support Vector Machine. SONAR dataset is used in this research work and the highest Mathews Correlation Coefficient of 0.72 is attained for Random Forest Classifier.

Keywords: Supervised Classification; DWT; KNN classifiers; SONAR dataset;

1.INTRODUCTION

Nowadays applications of Machine Learning (ML) are growing to a greater extent in various fields and some of the fields include risk management, fault diagnosis, decision control, disease diagnosis, etc. Supervised Machine Learning is a method of training the machine and the output is based on the trained dataset given in the past [1]. Fig. 1 depicts the general processes carried out in machine learning. Some of the popular supervised classification approaches include Linear Regression, K-Nearest Neighbours (KNN), Naive Bayes (NB), Random Forest Classifier (RFC), and Support Vector Machines (SVM), etc.

SONAR dataset is used in this research work. SONAR is a Spanish word meaning sound uses sound propagation technique to detect the objects at the signals off from it. In the mining industry, SONAR will be
used to identify the presence of minerals. For this purpose, a machine learning algorithm is required to classify the obtained data as either Mineral or Rock. This dataset contains 208 subjects obtained by sending SONAR Signals off at different angles and directions. Out of the 208 subjects, 111 subjects are obtained through reflection from Metal cylinder (minerals) and 97 subjects are obtained through reflection from Rocks. Every subject has 60 different reflections (features) values ranging from 0 to 1.

![Histogram of SONAR dataset features](image)

**Figure 2.** Histogram of SONAR dataset features

Histogram analysis will be an efficient tool to analyze the pattern of the features and to check whether the features are linearly separable or not. If the histogram has two distinct peaks and a valley, the features are linearly separable and the classification task is simple. The histogram of 60 features of all the 208 subjects available in the SONAR dataset is presented in Fig. 2. The histogram is skewed and only one peak is present and it clearly shows that the features are non-linearly separable.

The overall procedure followed is represented in Fig. 3. The SONAR dataset contains no missing values and the features vary in the range of [0, 1] already. Hence normalization techniques like min-max scaling are not required. The SONAR dataset features will be directly fed to anyone of the three data transformation techniques namely Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), and Hilbert Transform (HT). The transformed data will be fed to anyone of the four supervised classification techniques namely KNN, NB, SVM, and RFC.

![Overall Procedure followed in this research work](image)

**Figure 3.** Overall Procedure followed in this research work

The organization of remaining paper is given below: The basics of four different classifiers and three different transformation techniques used in this research work will be discussed in section 2. The experimental set-up and implementation procedure is elaborated in third section. The results are presented and discussed in fourth section and concluded in section 5.

2. **Fundamentals of Classifiers and Data Transformation Techniques:**

2.1 *Naive Bayes classifier*

Bayes theorem [2] is nothing but finding the probability of the selected event occurred to the given already occurred another event. Its represented as,

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$  (1)
Here in equation A and B refer to events. P(A) refers to prior probability P(A/B) denotes the posterior probability.

2.2 Random forest classifier

RFC [3] incorporates ensemble learning from decision trees. Mean Squared Error (MSE) will be used for solving regression problems in Random forest. As depicted in equation (2), this calculates the distance between each node from the predicted actual value.

\[ \text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \]  

Here in equation (2), N - total count of data points, \( \hat{y}_i \) – predicted value, \( y_i \) -original value.

2.3 K-nearest neighbor classifier

KNN algorithm [4] uses training data points and categorize the unknown data points using distance metric. Classification is based on the majority vote given by neighbors. A popular similarity metric, Euclidean distance from point p to point q is given by,

\[ D(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} \]  

2.4 Support Vector Machine

In the SVM classifier [5-6] after selecting the appropriate hyperplane, predictions will be made by using this hyperplane. The main function of this SVM classifier is to select a hyperplane which separates the given datapoints accurately. The SVM kernels are built through polynomials with order - p and Gaussian distribution as represented in equations (4) & (5) respectively.

\[ G(x_1, x_2) = (1 + x_1^T x_2)^p \]  
\[ G(x_1, x_2) = \exp(-\|x_1 - x_2\|^2) \]

2.5 Hilbert Transform

Hilbert transform [7] is used to obtain the minimum-phase response from a spectral analysis using the equation (6). Here \( u(t) \) is the input variable and \( H(u)(t) \) is the output variable.

\[ H(u)(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t-\tau} d\tau \]  

2.6 Principal Component Analysis

PCA [8] is used to transform correlated datapoints into uncorrelated data points. The main goal of PCA is to achieve a more compact model with lower dimensions of the dataset without losing any vital information.

2.7 Discrete wavelet transform

Discrete wavelet transform [9] returns the data vector as the same size of input data. The fact is that it decomposes into wavelets that are orthogonal to its translation and scaling.

\[ Z(\nu) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} x(t) \psi^\ast \left( \frac{t-a}{\alpha} \right) dt \]

Here \( \psi(\cdot) \) denotes the continuous mother wavelet, it is scaled by number \( \alpha \) and translated through number \( b \). There can be an infinite amount of wavelets.
3. IMPLEMENTATION METHODOLOGY

The SONAR dataset will be classified using four different classification algorithms and three different transforms to get the best estimator and appropriate best transform. The SONAR dataset will be divided into training and testing datasets using a stratified sampling technique which gives equal weightage to both classes (Minerals & Rocks). 75% of the subjects will be utilized for training whereas the remaining will be utilized for testing i.e., out of 208 subjects, 156 subjects are considered for training, and the remaining 52 subjects are considered for testing. During the training phase, grid search combined with K-Fold cross-validation (with K=10) is performed to avoid overfitting.

The possible values for parameters used in classifiers will be iterated in grid search and the ideal values for parameters are chosen based on the Balanced Accuracy score. These ideal parameters are used with the classifier to predict the test dataset. As an example of a grid search, the process for finding the ideal value of K and the ideal distance metric for the KNN classifier is depicted in Fig. 4. Here the grid values for K ranges between 1 to 15 and two different distance metrics namely Euclidean and Manhattan distance metrics are considered. In Fig. 4, the highest accuracy of 85% is attained when K = 7 with Manhattan distance. Hence to predict the test dataset KNN classifier will be used with K = 7 and manhattan distance.

![Fig. 4. Grid search for KNN classifier to find ideal distance metric and K value](image)

After predicting the class of the test dataset, the confusion matrix will be computed and six different performance metrics namely Mathews correlation coefficient (MCC), Accuracy, F1 score, Precision, Balanced Accuracy, and Error rate are computed for each classifier.

4. RESULTS AND DISCUSSION

The performance metrics of four different classifiers with three different transformation techniques for the classification of test datasets are shown in Table 1. Also, results obtained by individual classifiers are presented in Table 1 under the ‘No transform’ column. MCC can be considered as an efficient performance metric that gives an overall picture about the classification [10] and meaningful insights can be derived through MCC comparison for various classifiers and transforms as shown in Fig. 5.

![Fig. 5. MCC comparison of various classifiers with and without transforms](image)
Table 1. Performance metrics of various classifiers with and without transforms

| Transforms | No transform | DWT | HT | PCA | No transform | DWT | HT | PCA |
|------------|--------------|-----|----|-----|--------------|-----|----|-----|
| Classifiers/Metrics | Accuracy | Precision |
| KNN | 0.75 | 0.81 | 0.54 | 0.75 | 0.74 | 0.82 | 0.83 | 0.72 |
| RFC | 0.85 | 0.73 | 0.83 | 0.77 | 0.77 | 0.73 | 0.78 | 0.75 |
| NB | 0.63 | 0.67 | 0.62 | 0.63 | 0.73 | 0.73 | 0.66 | 0.73 |
| SVM | 0.85 | 0.75 | 0.81 | 0.65 | 0.79 | 0.77 | 0.78 | 0.66 |
| Classifiers/Metrics | Error rate | MCC |
| KNN | 0.25 | 0.19 | 0.46 | 0.25 | 0.50 | 0.61 | 0.21 | 0.50 |
| RFC | 0.17 | 0.27 | 0.17 | 0.23 | 0.72 | 0.46 | 0.66 | 0.54 |
| NB | 0.37 | 0.33 | 0.38 | 0.37 | 0.30 | 0.36 | 0.24 | 0.30 |
| SVM | 0.15 | 0.25 | 0.19 | 0.35 | 0.70 | 0.50 | 0.62 | 0.30 |
| Classifiers/Metrics | F1 Score | Balanced Accuracy |
| KNN | 0.77 | 0.82 | 0.24 | 0.78 | 0.74 | 0.81 | 0.57 | 0.74 |
| RFC | 0.87 | 0.75 | 0.85 | 0.80 | 0.83 | 0.73 | 0.86 | 0.76 |
| NB | 0.59 | 0.66 | 0.61 | 0.59 | 0.65 | 0.68 | 0.62 | 0.65 |
| SVM | 0.87 | 0.76 | 0.83 | 0.68 | 0.84 | 0.75 | 0.80 | 0.65 |

The highest MCC of 0.72 is attained for RFC without any transform. If PCA, HT, and DWT are used as transforms, then the MCC reduces. The same pattern is followed in the SVM classifier as well i.e., without transform, the MCC of the SVM classifier is 0.7 and the MCC reduces if transforms are used. But for KNN and NB classifiers, 22% and 20% accuracy increase is attained if DWT is used as transform when compared to ‘no transform’. HT is reducing the MCC and PCA is maintaining the MCC for both KNN and NB classifiers.

5. CONCLUSION

Investigations are done on the performance improvement of four popular supervised classification algorithms using three different transformation techniques and it is found that transformation techniques are good for some classifiers and not good for some other classifiers. For instance, DWT boosts the performance of KNN and NB classifier but reduces the performance of RFC and SVM. On the other hand, HT reduces the performance of all four classifiers. Hence it is essential to choose the appropriate transformation technique for the classifier and performance enhancement is also possible through the selection of appropriate transformation techniques.

References

[1] Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., Hinmikaiye, J.O., Olakanmi, O., Akinjobi, J., Supervised Machine Learning Algorithms: Classification and Comparison. International Journal of Computer Trends and Technology. 48(3), 128-138 (2017)

[2] Alberto Sanchis, Alfons Juan: A Word-Based Naïve Bayes Classifier for Confidence Estimation in Speech Recognition. IEEE Transactions on Audio, Speech, and Language Processing, 20(2), 565-564 (2012)

[3] Pal, M.: Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26(1), 217–222 (2005)
[4] Zhongheng Zhang: Introduction to machine learning: k-nearest neighbors. Ann. Transl. Med. 4(11), 1-7 (2016)

[5] Bharanidharan, N., Harikumar Rajaguru: Dementia MRI Classification Using Hybrid Dragonfly Based Support Vector Machine. 2019 IEEE R10 Humanitarian Technology Conference (R10-HTC), Depok, Indonesia, DOI: 10.1109/R10-HTC47129.2019.9042471 (2020)

[6] Benoit Magnin, Lilia Mesrob, Serge Kinkingnéhun, Mélanie Pélégrini-Issac: Support Vector Machine-based Classification of Alzheimer’s Disease from Whole-Brain Anatomical MRI. Neuroradiology, 51(2), 73-83 (2008)

[7] Olliffe I.T.: Principal Component Analysis, Series. Springer Series in Statistics (2002)

[8] King, Frederick, W.: Hilbert Transforms. Cambridge University Press, 1, (2010)

[9] Sundararajan, D.: Discrete Wavelet Transform: A Signal Processing Approach. Wiley. DOI: 10.1002/9781119113119 (2015)

[10] Chicco, Davide, Jurman, Giuseppe: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. DOI: 10.1186/s12864-019-6413-7 (2020)