Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2024.070108
ISSN 2371-8412 Vol. 7 Num. 1

The research on banknote authenticity discrimination analysis algorithm based on wavelet transform features

Zhao Sijian[1,a,*]

[1]*School of Statistics and Mathematics, Yunnan University of Finance and Economics, Wuhua District, Kunming City, Yunnan Province, 650221, China*
[a]*Zhao_Si-Jian@hotmail.com*
[*]*Corresponding author*

Keywords: Discriminant analysis, distance discriminant method, Fisher discriminant, Bayesian discriminant, authenticity identification

Abstract: In the authenticity identification of banknotes, features such as variance, skewness, kurtosis, and entropy of the images transformed by wavelet are used. This paper combines distance discriminant analysis, Fisher discriminant analysis, and Bayesian discriminant analysis for discrimination analysis. Variance can measure the texture complexity and grayscale level variation in the image, skewness is used to evaluate the symmetry and deviation from the normal distribution of the image, and kurtosis can measure the texture structure and grayscale level concentration. The entropy of the image reflects the complexity and uncertainty of the image. These features can be used to distinguish genuine banknotes from counterfeit ones.

1. Introduction

Background The authenticity identification of banknotes has always been an important research direction in the fields of finance and public security. With the development of technology, image processing and pattern recognition techniques have been widely applied in banknote authenticity identification. The classification analysis based on image features after wavelet transform and distance discrimination method, Fisher discrimination method, and Bayesian discrimination method have become hot research topics.

In terms of research background, the following aspects can be discussed:

1) The importance and application scenarios of banknote authenticity identification: The importance of banknote identification in the field of finance and public security cannot be underestimated, as well as its practical application scenarios in daily life, such as self-service vending machines, ATMs.

2) Current mainstream banknote authenticity identification methods: The main methods for authenticating banknotes are infrared detection and ultraviolet detection. Infrared detection scans banknotes using infrared radiation to identify their specific infrared spectral features. Ultraviolet detection involves illuminating banknotes and observing their fluorescent reactions under ultraviolet light to determine authenticity [1].

3) Application of image processing and pattern recognition in banknote authenticity

identification: The application of image processing and pattern recognition techniques in banknote authenticity identification includes image preprocessing, feature extraction, and classifier design.

4) Application of wavelet transform in image processing: Wavelet transform is a powerful mathematical tool, which is widely used in image processing. Its basic principle is to decompose the image into different frequency sub-bands by a set of wavelet functions [2].

5) Principles and applications of distance discrimination method, Fisher discrimination method, and Bayesian discrimination method: Distance discriminant method, Fischer discriminant method and Bayesian discriminant method are common pattern recognition methods, which play an important role in the identification of banknotes.

Through the explanation of the research background, readers can understand the importance of banknote authenticity identification, the limitations of existing methods, and the application value of wavelet transform and classification methods in this field. This information is instructive for the subsequent research content and method selection.

The article collects data by utilizing Kaggle's datasets, obtaining a total of 1098 data points. The collected data is then subjected to basic processing. Through the organized data, discriminant analysis, including distance discriminant analysis, Fisher discriminant analysis, and Bayesian discriminant analysis, is used to evaluate the classification effectiveness of banknotes. The results of the models are analyzed and summarized to draw conclusions. Based on the conclusions, appropriate targeted suggestions and solutions are presented.

2. Research objectives and significant

2.1 Objectives

The aim of this paper is to comprehensively apply distance discriminant analysis, Fisher discriminant analysis, and Bayesian discriminant analysis. By using the variance, skewness, kurtosis, and entropy of wavelet-transformed images as features, the study aims to discriminate between genuine and counterfeit banknotes. Analyzing these features allows quantification of the texture complexity, grayscale level variations, symmetry, deviation from normal distribution, texture structure, grayscale level aggregation, complexity, and uncertainty of banknote images. The research aims to enhance the accuracy and reliability of banknote authenticity identification, helping to prevent the circulation of counterfeit money and maintain financial security.

2.2 Significance

The research on banknote authenticity identification serves the purpose and significance of providing an efficient, accurate, and reliable method for distinguishing genuine banknotes from counterfeit ones. This is crucial for financial institutions, businesses, and individuals as the presence of counterfeit money can cause severe damage to economic transactions and social order.

By conducting research on banknote authenticity identification, the following objectives and significance can be achieved:

Maintaining financial security: Accurately determining banknote authenticity helps prevent the circulation of counterfeit money, safeguarding the security and stability of the financial system.

Preventing economic losses: Counterfeit money circulation leads to economic losses for businesses and individuals. Banknote authenticity identification technology can effectively reduce such losses.

Protecting consumer rights: Accurate identification of banknote authenticity ensures the protection of consumer interests, preventing them from falling victim to fraudulent activities during transactions.

Enhancing transaction efficiency: The application of banknote authenticity identification technology can expedite the transaction process, improve efficiency, and save time and labor costs.

In summary, the research on banknote authenticity identification holds significant importance in maintaining financial security, protecting consumer rights, preventing economic losses, and combating criminal activities. By selecting appropriate features and classification methods, combined with practical requirements and data characteristics, an efficient and reliable banknote authenticity identification model can be established, contributing to the socioeconomic development and stability of society.

3. Data description

Data Source: The data for this study is obtained from the Kaggle database, specifically from the following link: "https://www.kaggle.com/datasets/gauravduttakiit/banknote".

Variable Descriptions:

X1: VWTI - Variance of Wavelet Transformed Image

X2: SWTl - Skewness of Wavelet Transformed Image

X3: CWTI - Curtosis of Wavelet Transformed Image

X4: EI - Entropy of Image

Class: There are two classes, 0 and 1. Class 0 represents genuine banknotes and class 1 represents counterfeit banknotes.

4. Empirical analysis

Data Processing: First, the 1095 sample data were processed. Upon observation, counterfeit bank data it can be seen that the data has already been categorized in the table. The category "0" represents counterfeit bank data and the category "1" represents genuine banknotes. Among the samples, there are 607 classified as counterfeit bank data and 488 classified as genuine banknotes. We selected 607 counterfeit bank data data samples as class one, 290 genuine banknote data samples as class two, and 198 genuine banknote data samples for predicting classification effectiveness.

4.1 Model 1

Mahalanobis distance is a distance measurement method that takes into account the covariance structure between samples. It is used to measure the distance of a sample point relative to the mean vector of a given category in a multi-dimensional space [3].

Based on the theoretical basis mentioned above, the code follows these steps:

Check and process the input parameters. If the test sample TstG is not provided, the training samples classG1 and classG2 are combined as the test sample by default.

Calculate the number of rows in the test sample TstG and initialize the classification result matrix blong.

Calculate the mean vectors mu1 and mu2 for the training samples classG1 and classG2 respectively.

Based on the value of the parameter var.equal, calculate the discriminant function values w. If var.equal is TRUE or T, assuming equal variances, calculate the covariance matrix S for the combined samples. Use the Mahalanobis distance function to compute the Mahalanobis distance of the test sample TstG relative to mu2 and mu1, and subtract them to obtain the discriminant function value w.

If var.equal is FALSE, assuming unequal variances, calculate the covariance matrices S1 and S2 for class 1 and class 2 respectively. Use the Mahalanobis distance function to compute the Mahalanobis distance of the test sample TstG relative to mu2 and mu1, and subtract them to obtain the discriminant function value w.

Iterate through the discriminant function values w, classify the test samples as class 1 or class 2 based on certain conditions, and store the results in the classification result matrix blong.

Return the classification result matrix blong.

In summary, this code implements a discriminant analysis function based on the theoretical basis of discriminant analysis. It is used to classify and predict given training and test samples.

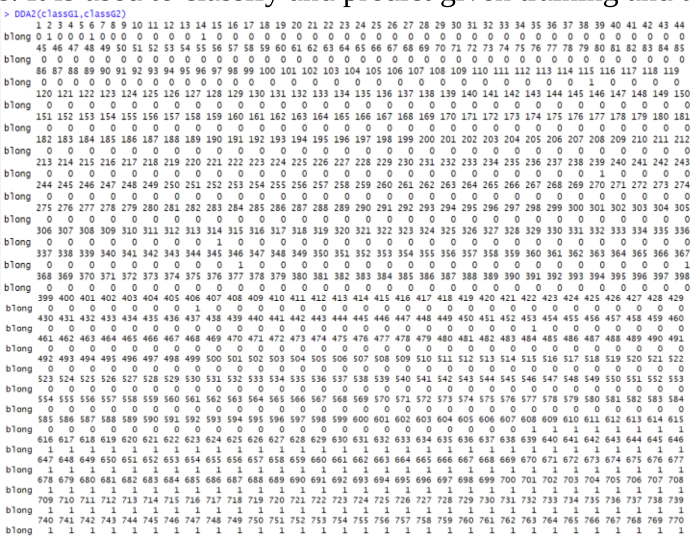

Figure 1: The results of classifying and predicting given training and test samples

The Figure 1 show that only counterfeit bills with serial numbers 2, 6, 14, 115, 406, and 453 were misclassified as genuine bills. It can be seen that out of 899 samples, only 6 were misclassified, resulting in a discrimination accuracy of 893/899 = 99.33%.

4.2 Model 2

The specific summary of the Fisher's discriminant method is as follows:

By using the attach (data) function, the variables in the dataset 'data' are loaded into the environment so that they can be directly operated using variable names.

The library (MASS) is used to load the MASS package in R, which provides functions for performing LDA.

A LDA model is created by calling the lda() function. The formula Class~X1+X2+X3+X4 specifies the model, indicating that the target variable is Class and the predictor variables are X1, X2, X3, and X4. Data [1:900,] represents using the first 900 samples in the dataset as training data.

The generated LDA model is assigned to the variable ld, so that it can be used for prediction or other operations later. The output result is as follows:

```
Call:
lda(Class ~ X1 + X2 + X3 + X4, data = data[1:900, ])

Prior probabilities of groups:
        0         1 
0.6755556 0.3244444 

Group means:
         X1        X2        X3        X4
0  2.322898  4.154896 0.8369718 -1.092019
1 -1.987436 -1.259352 2.5025157 -1.317595

Coefficients of linear discriminants:
            LD1
X1 -0.832242506
X2 -0.447299820
X3 -0.592249084
X4 -0.008778945
```

Figure 2: LDA model output results

In Figure 2, we have two classes labeled as 0 and 1. In the first 900 samples of the dataset, the prior probability of class 0 is 0.6755556, and the prior probability of class 1 is 0.3244444. The model also provides the mean vectors for each class. For class 0, the mean of X1 is 2.322898, the mean of X2 is 4.154896, the mean of X3 is 0.8369718, and the mean of X4 is -1.092019. For class 1, the mean of X1 is -1.987436, the mean of X2 is -1.259352, the mean of X3 is 2.5025157, and the mean of X4 is -1.317595. Lastly, the model provides the coefficients of the linear discriminant function. Based on these coefficients, a new data point can be classified using the linear discriminant function. For the linear discriminant function in this example, the coefficients for the discriminant variable LD1 are -0.832242506 (X1), -0.447299820 (X2), -0.592249084 (X3), and -0.008778945 (X4).

Then, the predict() function is used to classify the original data by backtesting it, comparing the output results of lda() with the true classification of the original data.

```
> cbind(Class[1:900],newG,Z$post,Z$x)
    newG            0            1         LD1
1  0    1 8.905620e-01 1.094380e-01  0.61081612
2  0    1 9.999999e-01 1.272475e-07 -2.13181812
3  0    2 1.553344e-02 9.844666e-01  1.85382072
4  0    1 9.999730e-01 2.704647e-05 -1.06522464
5  0    1 9.999909e-01 9.081888e-06 -1.28241457
6  0    1 1.000000e+00 2.925875e-08 -2.42436998
7  0    2 3.868687e-02 9.613131e-01  1.66747655
```

```
> tab=table(Class[1:900],newG)
> tab
   newG
      0   1
  0 585  23
  1   0 292
> sum(diag(prop.table(tab)))
[1] 0.9744444
```

Figure 3: Backtesting and comparison results

According to the Figure 3, it can be seen that 21 points, such as number 3 and number 7, were misclassified as counterfeit money. And the discrimination accuracy rate is 0.974444.

4.3 Model 3

The analysis method used is Linear Discriminant Analysis (LDA). LDA is a classic statistical classification method that maps high-dimensional data to a lower-dimensional space by finding the optimal projection direction, in order to achieve classification purposes. Specifically, the code uses the lda() function to fit an LDA model from the dataset. In the LDA model, the Class variable is treated as the response variable (or categorical variable), while X1, X2, X3, and X4 are treated as the predictor variables (or feature variables). By specifying the model formula Class~X1+X2+X3+X4, a linear classifier for predicting Class can be built. When fitting the LDA model, the code's data[1:900,] indicates that only the first 900 rows of the data dataset are used for modeling. This might be done to limit the number of training samples for better handling of computational load or to avoid overfitting. In summary, the code uses the LDA method to perform classification modeling on the given data and generates an LDA object (ld) that contains the model's parameters and results.

```
> library(MASS)
> ld = lda(Class~X1+X2+X3+X4,data = data[1:900,])
> ld
Call:
lda(Class ~ X1 + X2 + X3 + X4, data = data[1:900, ])

Prior probabilities of groups:
        0         1 
0.6755556 0.3244444 

Group means:
         X1        X2        X3        X4
0  2.322898  4.154896 0.8369718 -1.092019
1 -1.987436 -1.259352 2.5025157 -1.317595

Coefficients of linear discriminants:
            LD1
X1 -0.832242506
X2 -0.447299820
X3 -0.592249084
X4 -0.008778945
```

Figure 4: LDA that contains the model's parameters and results.

The Figure 4 displays some important information about the LDA model:

Prior probabilities: The prior probabilities of the predicted classification groups are as follows: The prior probability of class 0 is 0.6755556, and the prior probability of class 1 is 0.3244444.

Group means: The average values of each feature variable in each group are as follows: For class 0, the mean of X1 is 2.322898, the mean of X2 is 4.154896, the mean of X3 is 0.8369718, and the mean of X4 is -1.092019. For class 1, the mean of X1 is -1.987436, the mean of X2 is -1.259352, the mean of X3 is 2.5025157, and the mean of X4 is -1.317595.

Linear discriminant coefficients: These coefficients represent the contribution of each feature variable to the discriminant function in linear discriminant analysis. Here, only one discriminant function (LD1) is given, and the corresponding coefficients are as follows: The coefficient for X1 is -0.832242506, the coefficient for X2 is -0.447299820, the coefficient for X3 is -0.592249084, and the coefficient for X4 is -0.008778945.

```
> predict(ld,newdata = newdata)$class
  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [63] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[125] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[187] 1 1 1 1 1 1 1 1 1 1
Levels: 0 1
> tab = table(Class[1:900],newG)
> sum(diag(prop.table(tab)))
[1] 0.9744444
> 
```

Figure 5: Predicted regression results

According to the Figure 5, it can be seen that 21 points, including point 3 and point 7, were falsely classified as counterfeit bills. Additionally, the classification accuracy is 0.9744444.

5. Conclusion and recommendations

5.1 Conclusion

Based on the research on counterfeit banknote detection, this paper comprehensively utilizes distance discrimination, Fisher discrimination, and Bayesian discrimination methods, and analyzes the features of variance, skewness, kurtosis, and entropy of wavelet transformed images. The empirical results show that among these features, the Mahalanobis distance discrimination method performs better in counterfeit banknote detection.

Variance measures texture complexity and gray-level variation, skewness evaluates symmetry and deviation from normal distribution, and kurtosis measures texture structure and gray-level distribution aggregation. Entropy reflects complexity and uncertainty. Comprehensive use can distinguish genuine and counterfeit banknotes.

Although the aforementioned features are significant in counterfeit banknote detection, the empirical results demonstrate that the Mahalanobis distance discrimination method has better performance. However, the selection of the appropriate discrimination method should still be considered based on the specific application scenarios and dataset characteristics. Therefore, in practical applications, we should comprehensively assess the advantages and disadvantages of various discrimination methods and make reasonable choices to achieve more accurate and reliable counterfeit banknote detection.

5.2 Recommendations

Based on the empirical results, the Mahalanobis distance discrimination method performs better in counterfeit banknote detection. Therefore, in practical applications, it is recommended to prioritize the use of the Mahalanobis distance discrimination method for counterfeit banknote detection. Further research and improvement of counterfeit banknote detection methods can consider introducing more features, such as texture features and color features, to improve the accuracy of counterfeit detection. Additionally, the use of deep learning and other emerging technologies to detect counterfeit banknotes, can further improve the detection efficiency. In order to ensure the reliability and robustness of the detection system, it is necessary to expand and update the data set in order to adapt to the developing money manufacturing technology.

In conclusion, counterfeit banknote detection is an important field of application research, and combining different discrimination methods and utilizing various features can improve detection effectiveness. Future research should continue to focus on the development of this field, constantly improving and optimizing detection methods to meet society's needs for counterfeit banknote detection.

References

[1] Zhou, C. (2023). Study on the Wear Law of the Technical Characteristics of RMB Anti-Counterfeiting—Based on the Detection and Analysis of the Fifth Set of RMB 100 Banknotes in 2015 Edition. Journal of Tsinghua University (Science and Technology), 01-167-07.

[2] Men Xiuping. Application of Wavelet Transform in Image Processing [J]. Journal of Information Engineering, Anhui University of Finance and Economics, 2024, 2(19-24).

[3] Fei, Y., & Chen, Y. (2014). Multivariate statistical analysis. Beijing: Renmin University of China Press.

Appendix

```
library(readr)
data <- read_csv("C:/Users/Desktop/note.csv")
Class1 = data[2:609,1:4]
Class2 = data[610:900,1:4]
newdata = data[901:1096,1:4]
DDA2 <- function(class1, class2, YQ = NULL, var.equal = FALSE) {
  if (is.null(YQ)) YQ <- rbind(class1, class2)
  if (is.vector(YQ)) YQ <- t(as.matrix(YQ))
  else if (!is.matrix(YQ)) YQ <- as.matrix(YQ)
  if (!is.matrix(class1)) YQ1 <- as.matrix(class1)
  if (!is.matrix(class2)) YQ2 <- as.matrix(class2)
  nx <- nrow(YQ)
  blong <- matrix(rep(0, nx), nrow = 1, byrow = TRUE, dimnames = list("blong", 1:nx))
  mu1 <- colMeans(class1)
  mu2 <- colMeans(class2)
  if (var.equal == TRUE || var.equal == T) {
```

```
    S <- var(rbind(class1, class2))
    w <- mahalanobis(YQ, mu2, S) - mahalanobis(YQ, mu1, S)
  } else {
    S1 <- var(class1)
    S2 <- var(class2)
    w <- mahalanobis(YQ, mu2, S2) - mahalanobis(YQ, mu1, S1)
  }
  for (i in 1:nx) {
    if (!is.na(w[i]) && w[i] > 0) blong[i] <- 0
    else blong[i] <- 1
  }
  return(blong)
}
DDA2(class1,class2)
DDA2(class1,class2,newdata)
#fisher
attach(data)
library(MASS)
ld = lda(Class~X1+X2+X3+X4,data = data[1:900,])
ld
Z = predict(ld)
newG=Z$class
cbind(Class[1:900],newG,Z$post,Z$x)
output <- data.frame(Class = Class[1:900], newG, Z$post, Z$x)
write.csv(output, file = "output.csv", row.names = FALSE)
tab=table(Class[1:900],newG)
sum(diag(prop.table(tab)))
predict(ld,newdata = newdata)$class
predict(ld,newdata = newdata)$posterior
predict(ld,newdata = newdata)$x
# bayes
attach(data)
library(MASS)
ld = lda(Class~X1+X2+X3+X4,data = data[1:900,])
ld
Z = predict(ld)
newG = Z$class
cbind(Class[1:900],newG,Z$post,Z$x)
output <- data.frame(Class = Class[1:900], newG, Z$post, Z$x)
write.csv(output, file = "output.csv", row.names = FALSE)
predict(ld,newdata = newdata)$class
tab = table(Class[1:900],newG)
sum(diag(prop.table(tab)))
prenew = predict(ld,newdata = newdata)
cbind(prenew$class,prenew$post,prenew$x)
```