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) Currently the mainstream authenticity of the method: identify the authenticity of banknotes are mainly two methods of infrared detection. Infrared radiation scans the banknote to recognize the unique infrared spectral characteristics of the banknote. Ultraviolet detection through ultraviolet light irradiation of banknotes, observe its fluorescence in the ultraviolet light reaction, so as to determine the 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 that is widely used in image processing. The basic principle is to decompose the image into subbands of different frequencies by a set of wavelet functions [2].

5) Discriminant analysis works well for classification problems. Discriminant analysis is a statistical method mainly used for classification and pattern recognition problems. The core idea is to correctly classify a new sample into its own category by analyzing the features of different categories [3].

Through the explanation of the research background, the reader can understand the importance of identifying the authenticity of banknotes and the value of the application of the ripplelet transform and classification methods in this field by explaining the background of the study. It is understood that discriminant analysis is used in classification problems and identifying the authenticity of banknotes is also a binary classification problem. This information guides the selection of the content of the subsequent research and the selection of the research methodology.

The study compiles data sourced from Kaggle datasets, amassing a total of 1096 data points. Following data collection, initial processing procedures are executed. Evaluating the accuracy of banknote classification involves utilizing diverse methodologies like distance variance, Fisher's technique, and Bayesian discriminant analysis across the gathered dataset. The findings from these approaches are subjected to a lot of analysis and integration to provide final conclusions. Going by these facts, the recommendations and solutions given are accurately formulated.

2. Research objectives and significant

2.1 Objectives

This chapter intends to employ the distance judgment, the Fisher judgment, and the Bayesian judgment in order to increase the efficiency and the correctness of distinguishing the forged banknotes from the genuine ones. The objective is to determine the effectiveness of wavelet transformed images' features, including variability, asymmetry, shape, and randomness, to verify the genuineness of currency notes in order to select the most efficient method of verification.

This paper's primary objective is to improve the accuracy of banknote identification with the help of better techniques. Hence, through analysis of real and counterfeit currency, a fast and efficient mechanism to differentiate between the two has been established. This innovation, therefore, opens a new chapter in currency verification technology, and eliminates the occurrence of fake currency transactions in practical use thus enhancing the credibility of the financial systems.

In summary, this paper seeks to design efficient, precise and reliable systems for the identification of banknotes using different analytical methods and specific characteristics of the currency. These are important in enhancing the stability of financial systems and encouraging lasting development of the economy.

2.2 Significance

One of the main issues in today's society is the efficient identification of forged banknotes as soon as possible. The use of fake currency is a huge menace to the economy due to the fact that it threatens the safety and stability of monetary exchanges within the population.

Small amounts of counterfeit money need to be detected reliably so that it will not continue circulating within the financial system to compromise the financial system.

For the consumers to be safe against counterfeit notes, fake money has to be detected to promote the general well-being of the consumers, legit transactions and the confidence levels of the consumers.

To enhance the facets of market quality, then the process of identifying counterfeit notes has to be streamlined to improve market efficiency and encourage reliable and fair business interactions.

These measures not only minimize economic effects in terms of circulation of counterfeit notes, but also enhance the solidity of financial organizations as well as contribute to the growth of society's and the economy's wellbeing.

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 1 represents genuine banknotes and class 0 represents counterfeit banknotes.

4. Empirical analysis

Data Processing: First, the 1096 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 608 classified as counterfeit bank data and 488 classified as genuine banknotes. We selected 608 counterfeit bank data data samples as class one, 292 genuine banknote data samples as class two, and 196 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 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 and initialize the classification result matrix.

Calculate the mean vectors for the training samples respectively.

If var.equal is TRUE or T, the covariance matrix for the combined samples is computed assuming equal variances. Use the Mahalanobis distance function to compute the Mahalanobis distance of the test samples relative to the sample category mean vector, and then subtract the two to obtain the discriminant function value. If var.equal is FALSE, the covariance matrices of the two categories are computed separately, assuming unequal variances. The Mahalanobis distance function is used to calculate the Mahalanobis distance of the test sample with respect to the mean vector of the sample categories, and then they are subtracted to obtain the value of the discriminant function.

Iterate the discriminant function values, classify the test samples into 0 or 1 classes according to specific conditions and store the results in the classification result matrix.Return the classification result matrix.

In summary, these steps are based on the theoretical foundations of discriminant analysis and implement a discriminant analysis function. It is used to classify and predict given training and test samples.

```
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
blong 0 0 1 0 0 0 1 0 0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0
      28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
blong  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
      52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75
blong  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
      76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99
blong  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
      100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0
      118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0
      244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1
      316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
      334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0
      352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369
blong   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0
```

Figure 1. Partial presentation of classification results

The Figure 1 show that only counterfeit bills with serial numbers 3, 7, 15, 116, 240, 315, 347, 368, and the undisplayed 407, 454 were misclassified as genuine bills. It can be seen that out of 900 samples, only 10 were misclassified, resulting in a discrimination accuracy was approximately 98.89%. Finally, the prediction accuracy of the test set is obtained as 100% by running it through the R language

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:

Table 1: LDA model output results

	Groups	Prior Probabilities	X1	X2	X3	X4
	0	0.6755556	2.322898	4.154896	0.8369718	-1.092019
	1	0.3244444	-1.987436	-1.259352	2.5025157	-1.317595
coefficients			-0.832242506	-0.447299820	-0.592249084	0.008778945

In Table 1, 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, the coefficients for the discriminant variable LD1 are -0.832242506 (X1), -0.447299820 (X2), -0.592249084 (X3), and -0.008778945 (X4).

Table 2: Confusion Matrix

New G	Class_0	Class_1
0	585	23
1	0	292

Table 3: Test set prediction posterior probability

NO.	0	1
31	5.009199e-06	0.9999950
32	3.974186e-05	0.9999603
33	5.087928e-01	0.4912072
34	9.918076e-08	0.9999999

According to the Table 2, the Fisher discriminant yields a backward judgment prediction rate of 97.44%. According to the Table 3. The posterior probability of data point #33 is 0.4912072, which suggests that in the new data prediction, this data point is more likely to belong to categorization 0 than to categorization 1. So NO. 33 is misclassified, resulting in a discrimination prediction accuracy was approximately 99.49%.

4.3 Model 3

The third model is the naive Bayesian classifier, which is a probabilistic classification method based on Bayes' theorem and is often considered a type of Bayesian discriminant method [3]. The code uses naiveBayes function to fit a plain Bayesian model where Class variable is the response variable and X1, X2, X3 and X4 are the predictor variables. The model formula Class~X1+X2+X3+X4 is used to build the classifier. In modeling, the code uses only the first 900 rows of the dataset (data [1:900,]). The code classifies the data through a plain Bayesian approach and generates an object nb containing the model parameters and results.

Table 4: A priori probability distribution

Y	Probility
0	0.6755556
1	0.3244444

The Table 4 displays some important information about the prior probabilities of the classes in the dataset used to train the Naive Bayes classifier. Specifically, it shows the likelihood of an instance belonging to each class before considering any of the feature values.

From the table we can infer that the probability that an instance belongs to class 0 (Y = 0) is 67.56%, which indicates that class 0 is more common in the dataset. The probability that an instance belongs to class 1 (Y = 1) is 32.44%, which indicates that class 1 is less common in the dataset. For the Naive Bayes classification process, these probabilities are key starting points because they represent the base rate for each category, regardless of the effect of predictor variables. Understanding these baseline probabilities helps to properly understand the classifier output and make informed decisions based on the model predictions.

Table 5: Confusion Matrix

New G	Class_0	Class_1
0	552	56
1	78	214

According to Table 5, 552 instances with actual category 0 were correctly predicted to be 0 and 56 were incorrectly predicted to be 1. Of the instances with actual category 1, 214 were correctly predicted to be 1 and 78 were incorrectly predicted to be 0. Overall, the model's accuracy was 85.11%.Finally, according to the output of the program, 60 samples were misclassified, so the prediction accuracy is approximately 69.39%.

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.

The above features are significant in the detection of counterfeit banknotes. Empirical results show that the Mahalanobis distance discrimination method performs better than the other two methods. The selection of the appropriate discrimination method should be based on specific application scenarios, combined with the characteristics of the data set to consider. In practical application, the advantages and disadvantages of various discrimination methods should be comprehensively evaluated and reasonably selected, so as to achieve more reliable detection of counterfeit coins.

5.2 Recommendations

From the empirical results, the Mahalanobis distance discrimination method performs better in the detection of counterfeit banknotes. In practical applications, it is recommended to give priority to Mahalanobis distance discrimination method in detecting counterfeit banknotes. Further research to improve the counterfeit currency detection method, can consider the introduction of texture features and other features to improve the accuracy of counterfeit currency detection. The detection efficiency of counterfeit banknotes can be further improved by adopting emerging technologies such as deep learning. In order to ensure the reliability and robustness of the detection system to adapt to the evolving technology of banknote manufacturing, it is necessary to expand and update the data set.

Counterfeit banknote detection technology is a very major application research field, according to different detection methods and using different functions, it can enhance the detection efficiency of counterfeit banknote detection, which is a very major application field. Future scientific research will pay further attention to the development trend of this discipline and continue to innovate and optimise the detection technology to meet the needs of society for counterfeit currency detection technology.

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).

[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.