

Dependence in Classification of Aluminium Waste

Y Resti¹

Jurusan Matematika FMIPA Universitas Sriwijaya,
Jl. Raya Palembang-Prabumulih Km.32 Inderalaya 30662, Ogan Ilir, Sumatera
Selatan, Indonesia

E-mail: fyrest@yahoo.com

Abstract. Based on the dependence between edge and colour intensity of aluminium waste image, the aim of this paper is to classify the aluminium waste into three types; pure aluminium, not pure aluminium type-1 (mixed iron/lead) and not pure aluminium type 2 (un-recycle). Principal Component Analysis (PCA) was employed to reduction the dimension of image data, while Bayes' theorem with the Gaussian copula was applied to classification. The copula was employed to handle dependence between edge and colour intensity of aluminium waste image. The results showed that the classifier has been correctly classifiable by 88.33%.

1. Introduction

In an automatic sorting system, a classification of objects into a few groups mostly base on the images. An image usually represented by a vector in an $n.m$ dimensional space. In practice this spaces are too large to allow robust and fast classification so dimensionality reduction of data is needed at first. For classification, an object image allocated to a group base on a rule, for example a discriminant score or a posterior probability. Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are methods of classification was employed a grouping rule base on discriminant score. These methods assume that the joint probability density function is multivariate Gaussian distributed, also the marginal. However, this discriminant score cannot accurately classify object whose joint probability density function is different from its marginal distribution. A grouping rule using posterior probability base on Bayes' theorem is more flexible. In the theorem, the likelihood function can be nonparametric. If we have information about the image characteristics of each group, then we can use this information and the posterior probability based on Bayes' theorem in order to classify the objects. Dependence between the image characteristics can be also involved in this theorem by copula function.

PCA is a statistically techniques which is widely used for dimensionality reduction and recorded a great performance in many application both social and natural science. [1]-[7] applied Principal Component Analysis (PCA) and LDA in face recognition, while [15] applied PCA and QDA in classification of aluminium waste. This paper propose combination of PCA and Bayes' theorem with copula function to classify aluminium waste base on the images into a three groups; pure aluminium, not pure aluminium type-1 (mixed iron/lead) and not pure aluminium type-2 (un-recycle) involving edge and colour intensity of image. PCA as a technique to reduce the data dimension at first and Bayes' theorem with copula function as a technique to classification of aluminium waste image.

¹ Y Resti.

Copula was employed to handling the dependence between edge and colour intensity of the images. Using copula, joint distributions can be modelled curiosities of marginal distributions. Copula models have been applied in several areas such as finance [10]-[11], insurance [12]-[15], environmental [16] and classification studies [9], [17]-[18].

2. Materials and methods

The aluminum waste image is taken using a webcam by placing it on a conveyor belt with a particular pose (eg, front, top, back, down or side view), and then saved in jpeg format with a resolution of 320 x 240 pixels. For each of 85 different aluminum wastes, the images were taken in five pose; front view, top view, back view, down view and side view, so we collected 425 different images. Each image pose has been taken edge and colors component of RGB (Red, Green and Blue). The average of red, green and blue colors component is defined as color intensity.

Aluminum waste classification based on the dependence between edge and color intensity combine PCA and Bayes’ theorem with copula function. Principal Component Analysis (PCA) was employed to reduction the dimension of image data, while Bayes’ theorem with the Gaussian copula was applied to classification. The copula was employed to handle dependence between edge and color intensity of aluminum waste image.

1.1. Principal Component Analysis (PCA)

The objective of PCA approach is to get principal components which explain as much of the total variation in the data as possible with as few of these principle components as possible. The principal component is linear combination of observed variables.

For image data of aluminum waste, let observed variables are image poses, X_1, X_2, \dots, X_p . The linear combination of p image poses variables principal component is defined as,

$$Y_r = a_{r1}X_1 + a_{r2}X_2 + \dots + a_{rp}X_p = \mathbf{a}_r^T \mathbf{X}_k \tag{1}$$

where \mathbf{a}^T is eigenvector of covariance matrix \mathbf{S} which is orthogonal with each other. The eigenvector can be found by maximizing $\text{var}(Y_r) = \mathbf{a}_r^T \mathbf{S} \mathbf{a}_r$ with constraints $\mathbf{a}^T \mathbf{a} = 1$.

1.2. Copula in Bayes’s theorem

For the purpose of classification, the Bayes’ theorem stated in terms of density the following,

$$P(A = a|C = c) = \frac{P(C = c|A = a)P(A = a)}{P(C = c)} \tag{2}$$

where $P(A = a|C = c)$ is the posterior probability, $P(C = c|A = a)$ is the likelihood function, $P(A = a)$ is the prior probability and $P(C = c)$ is the data probability. An object image allocated to group g that have the image characteristic C given its attributes A if g has the largest posterior probability $P(A = a|C = c)$ between other group. For continuous attributes, a copula function can be used for modeling the dependence structure in the likelihood function.

Let $f_1(c_1|a)$ and $f_2(c_2|a)$ are the marginal density of image edge and colour intensity, respectively, while $F_1(c_1)$ and $F_2(c_2)$ are the marginal distribution successively. Let copula density function of the variables is $h(F_1(c_1), F_2(c_2)|a)$ and $P(A = a)$ is the prior probability based on the uniform distribution. Thus for this case, the copula in Bayes’ theorem such as [8] can be rewritten as,

$$P(A = a|C_1 = c_1, C_2 = c_2) = \frac{h(F_1(c_1), F_2(c_2)|a) \cdot f_1(c_1|a) \cdot f_2(c_2|a) \cdot P(A = a)}{f(c_1, c_2)} \tag{3}$$

For Gaussian bivariate copula, the density function $h(F_1(c_1), F_2(c_2)|a)$ is defined as,

$$h(\Phi(z_1), \Phi(z_2)) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\left(\frac{z_1^2 + 2\rho z_1 z_2 + z_2^2}{\sqrt{1-\rho^2}}\right)\right] \tag{4}$$

where $\Phi(z_1)$ and $\Phi(z_2)$ are the marginal standard Gaussian distribution of image edge and colour intensity, and ρ is parameter of Gaussian copula can be found by maximum likelihood estimation.

3. Results and Discussion

For each image edge, red, green and blue (RGB) colour component of image, the explanatory variables consists pixel size of down view (x_1), top view (x_2), side view (x_3), front view (x_4) and back view (x_5). We employed PCA in this research is to find as few of significant image poses variable as possible represented by principle components.

Table (1) shows eigenvalue, proportion of variance and cumulative proportion of principal component for image edge. The small eigenvalue shown the information of the image poses are useless so can be reduction without information loss. The cumulative proportion of the variance of the first component can explain 71% of the total variance and the second component when added to 82% means that if only taking one image pose alone is sufficient to represent the other four variables of image pose.

Table (1) Eigenvalue, proportion of variance and cumulative proportion of principal component for edge

Edge	Component				
	1	2	3	4	5
eigenvalue	4734.72	744.60	506.62	389.22	286.03
proportion of variance	0.71	0.11	0.08	0.06	0.00
cumulative proportion	0.71	0.82	0.90	0.96	1.00

Table (2) shows eigenvalue, proportion of variance and cumulative proportion of principal component for red, green and blue colour respectively. For red colour, the cumulative proportion of the variance of the first component explained 95% of the total variance. It means that if only taking one image pose alone is sufficient to represent the other four variables of image pose. For each of green and blue colour component, the cumulative proportion of the variance of the first component explained 96% and 94% respectively and it also means that if only taking one image pose alone is sufficient to represent the other four variables of image pose.

Table (2) Eigenvalue, proportion of variance and cumulative proportion of principal component for red, green and blue colour

Red	Component				
	1	2	3	4	5
eigenvalue	4747.22	170.80	60.72	26.12	21.30
proportion of variance	0.95	0.03	0.01	0.01	0.00
cumulative proportion	0.95	0.98	0.99	1.00	1.00

Green	Component				
	1	2	3	4	5
eigenvalue	6348.84	149.40	54.44	25.80	11.90
proportion of variance	0.96	0.02	0.01	0.00	0.00
cumulative proportion	0.96	0.98	0.99	1.00	1.00

Component

Blue	1	2	3	4	5
eigenvalue	4904.84	172.90	68.61	28.50	18.71
proportion of variance	0.94	0.03	0.01	0.01	0.00
cumulative proportion	0.94	0.97	0.98	0.99	1.00

The linear combination of image edge for the first principal component is

$$Y_{edge} = 0.456x_1 + 0.440x_2 + 0.492x_3 + 0.432x_4 + 0.412x_5$$

It means that the side view image pose has the most significant contribution to increasing of image edge comparing to the other image poses. Any increasing in one unit pixel of side view image pose will increase image edge at 0.492. Taking the side view image pose only has been representing the image edge. The linear combination of the first principal component for red, green and blue colour are given successively,

$$Y_{red} = 0.480x_1 + 0.493x_2 + 0.420x_3 + 0.421x_4 + 0.415x_5$$

$$Y_{green} = 0.481x_1 + 0.483x_2 + 0.422x_3 + 0.423x_4 + 0.422x_5$$

$$Y_{blue} = 0.486x_1 + 0.494x_2 + 0.419x_3 + 0.416x_4 + 0.415x_5.$$

It means that the top view image pose has the most significant contribution to increasing of red color composition comparing to the other image poses. Any increasing in one unit pixel of top view image pose will increase red color composition at 0.493. Taking the top view image pose only has been representing the red color composition. Similar as red colour, for both green and blue color component, the top view image poses of the components have the most significant contribution comparing to the other image poses. Any increasing in one unit pixel of each top view image pose will increase green and blue color composition successively at 0.483 and 0.494. Taking the top view image pose only of each color has been representing the green and blue color composition respectively. The PCA shown edge and red, green and blue colour has one significant image pose only, then the colour intensity is average of red, green and blue pixel in one pose only.

Testing 55 different images of aluminum waste to classify into 3 groups was employed. Table 3 and 4 shown results classification aluminum waste based on the image. Specially, the results in table 3 based on posterior probability as a classification rule, while in table 4 based on discriminant score such as done in [15] that employed QDA (Quadratic Discriminant Analysis). Employing Bayes' theorem with Gaussian copula such as given in (3) and (4) to classification 55 different images of aluminum wastes based on dependence between edge and colour intensity, we have been correctly classifiable by 88.33%. It is a performance improving of the method comparing with QDA that has been correctly classifiable by 80% as shown in Table 4.

Table 3. Result of Bayes Classification based on Gaussian Copula

Observed group	Result of Classification			% correct
	1	2	3	
1	25	0	0	100%
2	5	15	0	75%
3	0	1	9	90%
Total				88.33%

Table 4. Result of QDA Classification

Observed group	Result of Classification			% correct
	1	2	3	
1	25	0	0	100%
2	10	10	0	50%
3	0	1	9	90%
Total				80%

4. Conclusion

In this paper, we employed the combination of Principal Discriminant Analysis (PCA) and Bayes' theorem with Gaussian copula to classify the aluminium waste represented by the images, based on the dependence between edge and colour intensity of aluminium waste images. The combination of techniques has been correctly classifiable by 88.33% of 55 different images of aluminium wastes. It is a performance improving comparing with QDA (Quadratic Discriminant Analysis) that has been correctly classifiable by 80%. A better performance of classification can be applied to get a faster automatic sorting system.

5. References

- [1] W.S. Zheng, J.H. Lai and P.C. Yuen, "GA-Fisher: A New LDA-Based Face Recognition Algorithm with Selection of Principal Components", *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, Vol. 35, No. 5, 2005, pp. 1065-1078.
- [2] A. Eleyan and H. Demirel, "PCA and LDA Based Neural Network for Human Face Recognition" in *Face Reognition*, edited by K. Delac and M. Grgic, ISBN 978-3-902613-03-5, Vienna: I-Tech Education and Publishing, 2007, pp. 95 – 106.
- [3] N. Nain, P. Gour, N. Agarwal, R. P. Talawar and S. Chandra, "Face Recognition using PCA and LDA with Singular Value Decomposition(SVD) using 2DLDA", *Proceedings of the World Congress on Engineering*, Vol. I, London, U.K., July 2 - 4, 2008.
- [4] S. Singh, M. Sharma and N. S. Rao, "Accurate Face Recognition Using PCA and LDA", *International Conference on Emerging Trends in Computer and Image Processing*, Bangkok, 2011, pp. 62-68.
- [5] J. A. Dargham, A. Chekima, E. G. Moug, "Fusion of PCA and LDA Based Face Recognition System", *International Conference on Software and Computer Applications*, Singapore: IACSIT Press, vol. 41, 2012, pp. 111-116.
- [6] S.K. Hese and M.R. Banwaskar, "Performance of PCA and LDA for Face Recognition", *International Journal of Advanced Research in Electronics and Communication Engineering*, Vo. 2, No.2, 2013, pp. 149-154.
- [7] S. Fernandes and J. Bala, "Performance Analysis of PCA-based and LDA-based Algorithms for Face Recognition", *International Journal of Signal Processing Systems*, Vol.1 No.1, 2013, pp.1-6.
- [8] R.Salinas-Gutierrez, A.Hern'andez-Aguirre, M.J.J.Rivera-Meraz, and E.R.Villa-Diharce, Supervised Probabilistic Classification Based on Gaussian Copulas, *Advances in Soft Computing, Lecture Notes in Computer Science*, 2010, vol. 6438, pp 104-115.
- [9] C.M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.
- [10] D.E. Li, On default correlation: a copula function approach. *The Risk Metrics Group, Working Paper*, 2000, no. 99-07.

- [11] A. Shamiri., N.A. Hamzah and A. Pirmoradian, Tail Dependence Estimate in Financial Market Risk Management: Clayton-Gumbel Copula Approach, *Sains Malaysiana*, 2011, vol. 40 no. 8, pp. 927–935.
- [12] E.W. Frees, and P. Wang, Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics*, 2006, vol. 38, pp. 360-373.
- [13] Y. Resti, N. Ismail and S.H. Jaaman, Handling the dependence of claim severities with copula models, *Journal of Mathematics and Statistics*, 2010, vol. 6 no. 2, pp. 136-142.
- [14] Y. Resti, N. Ismail and S.H. Jaaman, Mathematical modeling for claim severities using normal and t copulas. *International Journal of Applied Mathematics and Statistics* 2012, vol. 27 no.3, pp. 8-19
- [15] Y. Resti, Statistically Classification of Aluminum Waste, *Proceeding of the 1st International Conference on Education, Technology and Science*, Jambi, Indonesia, November 19-20, 2014, pp. 414-417.
- [16] G. Wong, M.F. Lambert and A.V. Metcalfe, Trivariate copulas for characterization of droughts, *Australian and New Zealand Industrial and Applied Mathematics Journal*, 2008, vol. 49, pp 306-323.
- [17] Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer (2009)
- [18] Mackay, D.J.C.: *Information Theory, Inference, and Learning Algorithms*. Cambridge (2008)