머신 러닝을 사용한 이미지 클러스터링: K-means 방법을 사용한 InceptionV3 연구

닌담 솜사우트, 이효종
전북대학교 공과대학
nindam.somsawut@jbnu.ac.kr, hlee@jbnu.ac.kr

Image Clustering Using Machine Learning : Study of InceptionV3 with K-means Methods.

Somsauwt Nindam, Hyo Jong Lee*
Division of Computer Science and Engineering, Jeonbuk National University
*Corresponding author

Abstract

In this paper, we study image clustering without labeling using machine learning techniques. We proposed an unsupervised machine learning technique to design an image clustering model that automatically categorizes images into groups. Our experiment focused on inception convolutional neural networks (inception V3) with k-mean methods to cluster images. For this, we collect the public datasets containing Food-K5, Flowers, Handwritten Digit, Cats-dogs, and our dataset Rice Germination, and the owner dataset Palm print. Our experiment can expand into three-part; First, format all the images to un-label and move to whole datasets. Second, load dataset into the inception V3 extraction image features and transferred to the k-mean cluster group hold on six classes. Lastly, evaluate modeling accuracy using the confusion matrix base on precision, recall, F1 to analyze. In this our methods, we can get the results as 1) Handwritten Digit (precision = 1.000, recall = 1.000, F1 = 1.00), 2) Food-K5 (precision = 0.975, recall = 0.945, F1 = 0.96), 3) Palm print (precision = 1.000, recall = 0.999, F1 = 1.00), 4) Cats-dogs (precision = 0.997, recall = 0.475, F1 = 0.64), 5) Flowers (precision = 0.610, recall = 0.982, F1 = 0.75), and our dataset 6) Rice Germination (precision = 0.997, recall = 0.943, F1 = 0.97). Our experiment showed that modeling could get an accuracy rate of 0.8908; the outcomes state that the proposed model is strongest enough to differentiate the different images and classify them into clusters.

1. Introduction

Nowadays, artificial intelligence (AI) is becoming more powerful and highly efficient than in the past [1]. Machine learning (ML) and deep learning (DL) are subsections of AI and are very popular in the areas. Many fields using AI, such as industries, medical, agricultural, educational, and so on. AI is used to improve the productivity and to reduce the cost. These are the parts mentioned assume to study AI technology.

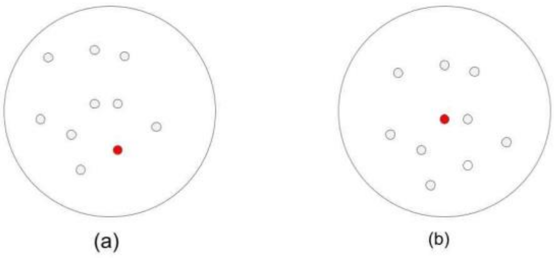
This research applies unsupervised learning methods to cluster images without labeling. To begin with, we collected the public datasets that are Handwritten Digit, Food-K5, Cats-dogs, Flowers, [2], owner Palm print dataset [3], and our dataset Rice Germination [4] of different images. Each dataset contains images with different colors, textures, and shape features without labeling [5]. Then we applied the Inception V3 with the K-means method [6] [7] [8] to cluster images. Finally, the model is evaluated, and the performance is measured regarding the precision, recall, and F1 to measure the accuracy using a confusion matrix.

The remaining of the paper is organized as follows. Section 2-3 reviews the related works and methods used in clustering methods, discusses the proposed unsupervised learning method in image clustering. The performance evaluation is analyzed in Section 4. The concluding remarks are given in Section 5.

2. Preliminaries

2.1 K-means Algorithm

Most of the clustering techniques proposed are simple methods of clustering. The methods choose either K data points in the datasets either manually or randomly partition the data points into K subsets. More sophisticated methods such as density-based initialization, intelligent initialization, and farthest first initialization pick the center point randomly and add more center points next most distant from the existing ones. The K-means clustering method partitions the data points into K clusters. The example of K-means clustering with K = 1 on a set of points is illustrated in Figure 1. The clusters are initialized by randomly selecting two points as centers [9].



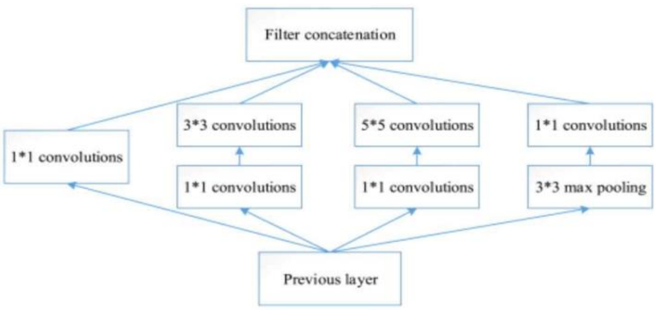
(Figure 1) (a) Initialization, (b) Re-assignment.

The objective function of K-means is to minimize the sum of pairwise Euclidean distances of the points in clusters. The steps of the clustering process are as follows:

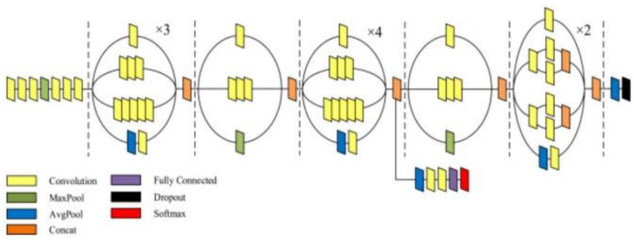
- ✧ Choose K centroids randomly.
- ✧ Calculate the Euclidean distance of each point to the centroids.
- ✧ Assign each point to the closest centroid.
- ✧ Update each cluster center by the mean distance of all points in the cluster.
- ✧ Repeat the second and third steps until they converge [10] all points.

2.2 Inception v3

The GoogLeNet network is a CNN that was developed by Google teams in 2014. There are adopt the Inception network structure in a new way to reduces the amount of network parameters and increases the network depth. Culminate in this architecture is widely used in image classification tasks. As a set of the GoogLeNet network is the Inception network structure, the GoogLeNet network is called the Inception network. There are many versions of GoogLeNet, separate the groups as Inception v1 (2014), Inception v2 (2015), Inception v3 (2015), Inception v4 (2016), and Inception-ResNet (2016). the Inception architecture is a module that typically has three types different sizes of convolution and one maximum pooling. In the previous layer for the network output, the channel is aggregated after the convolution operation, after that the nonlinear fusion is performed. This architecture presents the expression of the network and the adaptability to different scales can be improved, and over-fitting can be prevented. the inception network structure is shown in figure 2 and the example of Inception v3 network architecture is shown in figure 3 [11].



(Figure 2) Inception network structure [11].



(Figure 3) Inception V3 network structure [11].

2.3 Evaluation Metrics

In this experiment, we used the confusion matrix to evaluate the performance of the proposed model. In the evaluation, Precision, Recall, and F-measure metrics were considered. The confusion matrix is a two-dimensional array that contains the results of actual and predicted classes and explains how the clustering of new samples is performed. The confusion matrix can be used for multi-class clustering. Figure 4 shows an example of a multi-class confusion matrix where the attributes $\{C_1, \ldots, C_n\}$ are the number of samples that are classified as class C_j but belong to class C_i. The best clustering will have zero value in non-diagonal entries. the confusion matrix is shown in Figure 4.

	Predicted				
Actual		C_1	C_2	...	C_n
	C_1	N_{11}	N_{12}	...	N_{1n}
	C_2	N_{21}	N_{22}	...	N_{2n}
	⋮	⋮	⋮	⋮	⋮
	C_n	N_{n1}	N_{n2}	...	N_{nm}

(Figure 4) Confusion matrix.

Single-class metrics are evaluated for every class and are less sophisticated to class imbalance. Therefore, it is more suited to evaluate clusters in skew data domains. The single-class metrics are detailed as below:

Precision metric measures correctly classified positive class samples using Eq. (1). Recall measures the percentage of identified all actual positive samples derived in Eq. (2). And F-measures derived in Eq. (3). are used to analyze the intuitive trade-off between precision and recall so that the constraint can be adjusted.

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$F-measure = \frac{(1+\beta)PrecisionRecall}{\beta^2 Precision + Recall} \tag{3}$$

Multi-class can be computed globally over all classes, and its elements (Precision and Recall) are obtained by summing over all classes. While macro-average can be obtained from each class locally and averaging those, a single value can be computed as an F-measure value in the whole classes [12].

(Figure 5) Sample dataset.

3. Proposed Method

3.1 Collect the dataset

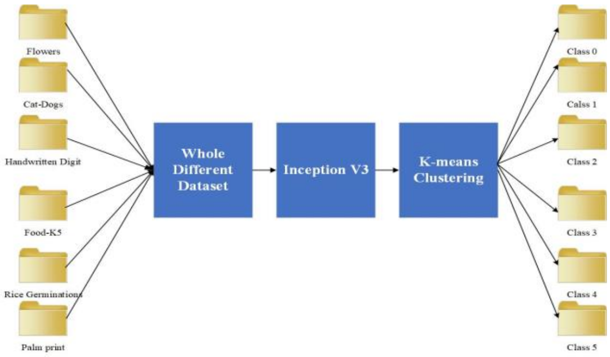
We collect the public dataset consisting of four public datasets and one owner dataset and one of our datasets, enumerate that is respectively Handwritten Digit, Food-K5 (foods), Palm print, Cats-Dogs, Flowers, and Rice Geminations. We randomly 1,250 images shown in table 1. and visualize the whole dataset shown in figure 5.

<Table 1> Collection datasets.

No.	Name	Amount/Class	Amount/images
1	Handwritten Digit	10	1,250
2	Food-K5 (foods)	1	1,250
3	Palm print	100	1,250
4	Cats-Dogs	2	1,250
5	Flowers	5	1,250
6	Rice Geminations	2	1,250
Total			7,500

3.2 Overall our methods

We randomly selected 1,250 images in each class and then combined all of the images composed into a whole dataset. After that, we converted it into unification labeling, then loaded whole datasets into the inception V3 architecture extraction feature priors feed to the K-means clustering. In this case, we assign the K cluster to six classes, and overall, our methods are shown in figure 6.



(Figure 6) Procedure pipeline.

4. Experiments

In this experiment, the inception V3 architecture with the K-means method can cluster the images into class, as shown in Figures 7-9.

(a) (b)

(Figure 7) (a) Class Handwritten Digit and (b) Class Food-K5.

(c) (d)

(Figure 8) (c) Class Palm print and (d) Class Cats-Dogs.

(e) (f)

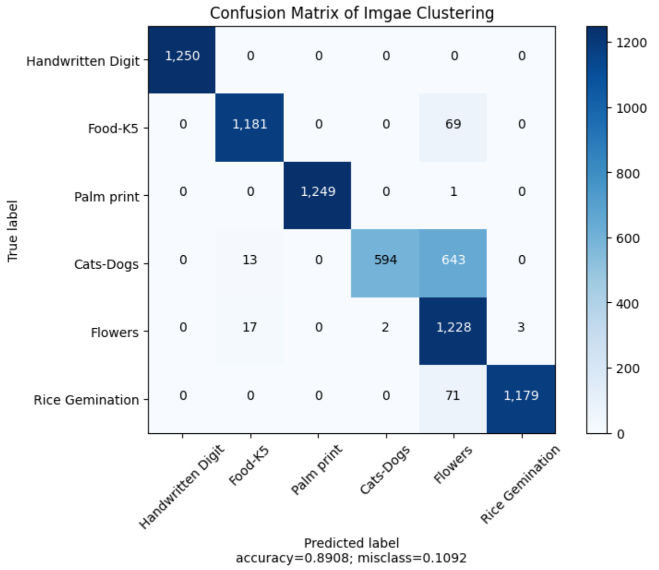
(Figure 9) (e) Class Flowers and (f) Class Rice Geminations.

This step measures the performance of the model using the precision, recall, and F1 or F-measure in the confusion matrix to analyze the modeling. Comparison of each class accuracy between class, Handwritten Digit, Food-K5 (foods), Palm print, Cats-Dogs, Flowers, and Rice Geminations is shown in Table 2.

<Table 2> The results of precision, recall, and F1.

	Datasets	Class-0	Class-1	Class-2	Class-3	Class-4	Class-5	Precision	Recall	F1 - Score
0	Handwritten Digit	1250	0	0	0	0	0	1.000	1.000	1.00
1	Food-K5	0	1181	0	0	69	0	0.975	0.945	0.96
2	Palm print	0	0	1249	0	1	0	1.000	0.999	1.00
3	Cats-Dogs	0	13	0	594	643	0	0.997	0.475	0.64
4	Flowers	0	17	0	2	1228	3	0.610	0.982	0.75
5	Rice Gemination	0	0	0	0	71	1179	0.997	0.943	0.97

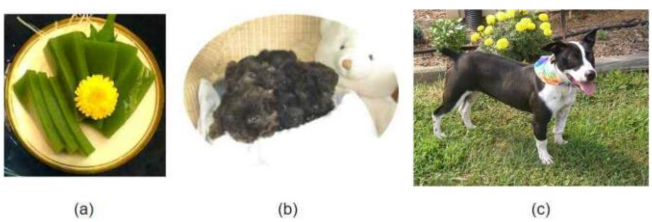
After we can get the result mentioned above, each class can use the confusion matrix analysis of the accuracy of all dataset, the confusion matrix is shown in Figure 10.



(Figure 10) Confusion Matrix Measures.

In this our methods, we can get the results as 1) Handwritten Digit (precision = 1.000, recall = 1.000, F1 = 1.00), 2) Food-K5 (precision = 0.975, recall = 0.945, F1 = 0.96), 3) Palm print (precision = 1.000, recall = 0.999, F1 = 1.00), 4) Cats-dogs (precision = 0.997, recall = 0.475, F1 = 0.64), 5) Flowers (precision = 0.610, recall = 0.982, F1 = 0.75), and our dataset 6) Rice Germination (precision = 0.997, recall = 0.943, F1 = 0.97). The result of our experiment showed that modeling can get the accuracy rate of 0.8908.

However, the modeling mistake is 0.1092 to clustering true class, limited to decision and classification to the actual class., Some example of failure cases is shown Figure 11.



(Figure 11) Failure images in the class.

5. Conclusion

In conclusion, we proposed a model which can differentiate different images and cluster similar images into the same class with an accuracy of 0.8908. However, in the case of Cats-Dogs grouping, the Inception V3 construction with K-means clustering was not successful. Hence, in the next experiment, we need to improve the modeling with new methods to increase accuracy, such as using VGG16 or another deep neural network architecture with K-means methods to enhance its clustering accuracy efficiency.

Acknowledgment

This research was supported by Basic Science Research Program through the NRF of Korea funded by the Ministry of Education (GR 2019R1D1A3A03103736).

Reference

[1] Synced G, Shaoyou (Victor) L, Baorui (Alex) C, Qingyan T, Chenchen (Chain) Z, Chen (Robert) T, Meghan H. (2018). Year of AI: How Did Global Public Company Adapt to the Wave of AI Transformation: A 2018 Report about Fortune Global 500 Public Company Artificial Intelligence Adaptivity (Kindle Edition). ASIN: B07K91RKVK.

[2] kaggle. (2021). Datasets. https://www.kaggle.com/datasets.

[3] Aung Si Min Htet, & Hyo Jong Lee. (2021). TripletGAN VeinNet: Palm Vein Recognition Based on Generative Adversarial Network and Triplet Loss. ICCEAI 2021, Shanghai, China.

[4] Nindam, S., Sung, T. L., Manmai, T. O., & Lee, H. J.(2019). Collection and Classification of Jasmine Rice Germination Using Convolutional Neural Networks, International Symposium on Information Technology Convergence (ISITC2019), Tianjin, China.

[5] Sleit, A., Abu Dalhoum, A. L., Qatawneh, M., Al-Sharief, M., Al-Jabaly, R., & Karajeh, O. (2011). Image clustering using color, texture and shape features. KSII Transactions on Internet and Information Systems (TIIS), 5(1), 211-227.

[6] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[7] Cohn, R., & Holm, E. (2021). Unsupervised machine learning via transfer learning and k-means clustering to classify materials image data. Integrating Materials and Manufacturing Innovation, 1-14.

[8] Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M., & Van Gool, L. (2020, August). Scan: Learning to classify images without labels. In European Conference on Computer Vision (pp. 268-285). Springer, Cham.

[9] Shojanazeri, H., Aryal, S., Teng, S. W., Zhang, D., & Lu, G. (2018, November). Image clustering using a similarity measure incorporating human perception. In 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

[10] Jin X., Han J. (2011) K-Means Clustering. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_425.

[11] Dong, N., Zhao, L., Wu, C. H., & Chang, J. F. (2020). Inception v3 based cervical cell classification combined with artificially extracted features. Applied Soft Computing, 93, 106311.

[12] Tanha, J., Abdi, Y., Samadi, N. et al. Boosting methods for multi-class imbalanced data classification: an experimental review. J Big Data 7, 70 (2020).