Unsupervised Image Classification Using Multi-Autoencoder and K-means++

Shingo Mabu
Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Tokiwadai2-16-1
Ube, Yamaguchi 755-8611, Japan

Kyoichiro Kobayashi
Department of Information Science and Engineering, Faculty of Engineering, Yamaguchi University, Tokiwadai2-16-1
Ube, Yamaguchi 755-8611, Japan

Masanao Obayashi
Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Tokiwadai2-16-1
Ube, Yamaguchi 755-8611, Japan

Takashi Kuremoto
Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Tokiwadai2-16-1
Ube, Yamaguchi 755-8611, Japan

E-mail: mabu@yamaguchi-u.ac.jp, m.obayas@yamaguchi-u.ac.jp, wu@yamaguchi-u.ac.jp

Abstract

Supervised learning algorithms such as deep neural networks have been actively applied to various problems. However, in image classification problem, for example, supervised learning needs a large number of data with correct labels. In fact, the cost of giving correct labels to the training data is large; therefore, this paper proposes an unsupervised image classification system with Multi-Autoencoder and K-means++ and evaluates its performance using benchmark image datasets.

Keywords: neural network, deep autoencoder, K-means++, clustering

1. Introduction

Recently, deep learning[1] has been actively applied to various problems and shown distinguished performance. Especially, in image classification problems, since deep neural networks can automatically extract features contained in each image[2], it becomes easier to apply deep neural networks to image classification comparing to the conventional methods that extract features using manually created feature extraction method.

However, classification systems using deep learning are generally based on supervised learning. Thus, a large number of training data with correct labels are necessary for the learning. Since it is a tough task to give correct labels to many (hundreds ~thousands) images, this paper proposes an unsupervised image classification (clustering) algorithm that does not use data with correct labels to reduce the cost of preparing training data. However, when we do not use correct labels, the classifier cannot know the explicit information on the classes. Therefore, sufficient information to distinguish the differences between the images is necessary for accurate classification. To obtain sufficient information, we apply some image processing techniques to increase the variation of the images. Then, the increased images are inputted to Multi-autoencoder to extract feature information. In summary, the proposed method has two features; 1) increasing the variation of original images using some image processing techniques, and 2) extracting features using Multi-autoencoder that uses the original and generated images. Autoencoder-based data clustering has been already proposed[3]. However, the aim of the conventional method is to enhance Single-autoencoder, while the aim of the proposed method is to evaluate the effectiveness of Multi-autoencoder.

The rest of this paper is organized as follows. In section 2, the proposed unsupervised image classification algorithm is explained. In section 3, a benchmark image dataset (caltech101) and experimental settings are explained, then experimental results are shown. Section 4 is devoted to conclusions.

2. Unsupervised image classification algorithm

The flow of the proposed method is shown in Fig. 1. First, some image processing techniques (gray-scale transformation and edge detection by Sobel filter[4]) are applied to the original images to generate variations. Second, several independent autoencoders (Multi-autoencoder) are prepared (the number of autoencoders is three corresponding to the number of datasets, i.e., the original dataset and two generated datasets). Then, feature values are obtained by the Multi-autoencoder. Finally, the image clustering is carried out by K-means++ algorithm.

2.1. *Feature extraction using Image processing and Multi-autoencoder*

The image dataset used in this paper is caltech101[5] that is a set of color natural images (32×32 pixel) such as watch, motorbike, airplane, grand piano, etc. We apply gray-scale transformation and Sobel filter independently to the original dataset, and two new datasets are generated (Fig. 2). Then, the original and generated images are encoded by Multi-autoencoder. Fig. 3 shows 9-layered autoencoder used for the feature extraction. The original or generated images inputted to the autoencoder are encoded by second–fifth layers, decoded by sixth–ninth layers and finally reconstructed images can be obtained. The autoencoder is trained so that the mean squared error between the input and output values is minimized. After finishing the training, the feature values of the input image are obtained by the fourth layer.

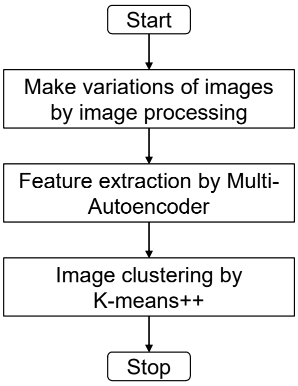
Start
Make variations of images by image processing
Feature extraction by Multi-Autoencoder
Image clustering by K-means++
Stop

Fig. 1. Flowchart of the proposed method

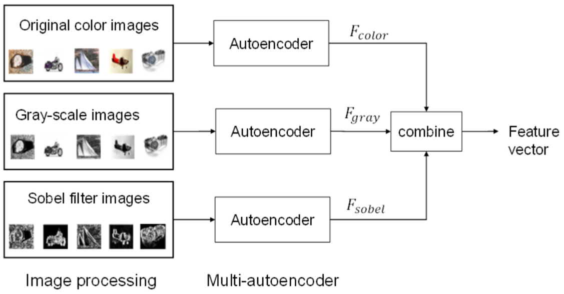

Fig. 2. Feature extraction using image processing and Multi-autoencoder

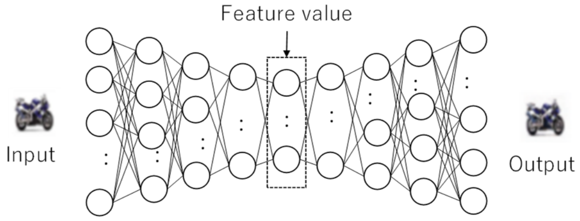

Fig. 3. Structure of autoencoder

Here, let the original image data number be *d*, the feature values of data *d* extracted by the autoencoder for the color image be $F_{color}(d)$, those for the gray-scale image be $F_{gray}(d)$, and those for the Sobel filter image be $F_{sobel}(d)$. Then, a feature vector combining $\{F_{color}(d), F_{gray}(d), F_{sobel}(d)\}$, $\{F_{color}(d), F_{gray}(d)\}$, $\{F_{color}(d), F_{sobel}(d)\}$ or $\{F_{sobel}(d), F_{gray}(d)\}$ is defined by $F(d)$ which is used in the next K-means++ clustering[6].

2.2. *Clustering by K-means++*

K-means[7] is a clustering algorithm that makes groups (clusters) of data where each data belongs to one of the groups whose centroid is the nearest. K-means++ is an extension of K-means that improves the initialization of cluster centroids. The centroid initialization method of K-means++ is summarized as follows.

1. Randomly select one data *d* out of *N* data as a cluster centroid. Here, let *m* be 1.
2. Repeat the following until *m*=*K* (*K* is the total number of clusters).
   i. Calculate $D(d)$ which is the Euclidian distance between data d and the nearest centroid.
   ii. Select one data $d(d = 1,2,\dots,N-m)$ with probability $\phi(d)$ as a new centroid.
   $$\phi(d) = \frac{D(d)^2}{\sum_{d=1}^{N-m} D(d)^2}$$
   iii. $m \leftarrow m + 1$

By applying K-means++ to all the feature vectors $F(d)$, similar images are put into the same cluster.

3. Experiment

The image data set is caltech101 which contains pictures of objects belonging to 101 classes. In this paper, images of 10 classes are selected and used for the experiment. The 10 classes are airplanes, bonsai, faces_easy, hawksbill, ketch, leopards, motorbikes, watch, grand_piano, chandelier, and each class contains 90 images (totally 900 images). The clustering performance of the proposed method is evaluated by F-measure comparing to the conventional methods. The conventional methods include 1) clustering for the features extracted by single-autoencoder, 2) clustering for the original images without feature extraction, 3) clustering for the Histogram of oriented gradients (HOG) features[8] of the original images. The parameters of Multi-autoencoder are shown in Table 1. The number of layers is nine, and the number of input nodes is determined according to the sizes of input images. The size of the original color images is 3072 (32 pixel × 32 pixel × 3 channels) and that of gray-scale images and Sobel filter images is 4096 (64 pixel × 64 pixel × 1 channel). Since gray-scale and Sobel filter images have one channel, the size is enlarged to increase the number of pixel information. The number of nodes in the 5th layer is 256 or 171 depending on the combination of $\{F_{color}(d), F_{gray}(d), F_{sobel}(d)\}$. When two kinds of feature values are combined to make $F(d)$, 256 nodes are used and obtain $F(d)$ with 512 ($= 2 \times 256$) values. When all the three kinds of feature values are combined, 171 nodes are used to obtain $F(d)$ with 513 ($= 3 \times 171$) values. This is for making the dimensional size of feature vectors be almost the same between the different settings. Multi-autoencoder is learned by stochastic gradient descent method with Adam (adaptive moment estimation)[9] which is an online estimation method of appropriate learning rate.

Table 1. Parameters of Multi-autoencoder

# of nodes (1st layer–5th layer)	3072 or 4096, 2048, 1024, 512, 256 or 171
Activation function	Sigmoid
Pre-training epoch	300
Fine-tuning epoch	200
Learning algorithm	Adam

Table 2 shows the clustering performance of the proposed method and some conventional methods averaged over 10 independent trials. The proposed

Table 2. Clustering performance

Method	Image set	Precision [%]	Recall [%]	F-measure [%]
Proposed method with Multi-autoencoder	Color+Sobel	73.6	69.1	71.3
	Color+Gray	66.6	62.4	64.4
	Gray+Sobel	69.6	67.7	68.6
	Color+Gray+Sobel	68.2	65.4	66.7
Conventional method with Single-autoencoder	Color	69.4	64.9	67.1
	Gray	62.9	61.8	62.6
	Sobel	68.5	65.6	67.0
Conventional method without autoencoder	Color	60.8	57.6	59.1
	HOG feature	69.8	62.7	66.0

Table 3. Results of t-test

Methods	p-value
Color+Sobel vs. Color	1.23×10^{-4}
Color+Sobel vs. Sobel	7.62×10^{-4}

Color+Sobel: Proposed method with color and Sobel filter images
Color: Conventional method (Single-autoencoder) with color images
Sobel: Conventional method (Single-autoencoder) with Sobel filter images

method with color and Sobel filter images shows the best F-measure of 71.3%. The proposed method with gray and Sobel filter images shows the second best F-measure of 68.6%. From these results, we can see that the Sobel filter images can give additional useful information for the clustering. The proposed method with all the three kinds of images does not show the best result because too much (redundant) information would even cause bad effect on the learning efficiency. Among the three settings of Single-autoencoder, the method with color images shows the best F-measure; thus, we can say that the color original image contains the useful information comparing to gray and Sobel filter images. However, the difference of F-measure between color and Sobel filter is small; thus, we can see the Sobel filter still contains useful information. As for the method without autoencoder, HOG feature shows better F-measure than the method using raw color images.

Table 3 shows the results of t-test for F-measure between the proposed method with color and Sobel filter images and the conventional methods with color or Sobel filter images. We selected the methods for the t-test that show better F-measure among all the settings of the proposed method or conventional methods. From Table 3, we can see that there are significant differences between the proposed method and conventional methods.

Fig. 4 shows an example of the generated cluster. This cluster mainly contains grand_piano images; however, some other classes of images whose features are very similar are also contained. The squares in Fig. 4 show the images that are not grand_piano. The proposed method is based completely on unsupervised learning, thus it is difficult to realize perfect clustering due to the lack of class information. Nevertheless, the proposed method can generally grasp the essential features for selecting similar images.

4. Conclusions

This paper proposed an unsupervised image classification algorithm using Multi-autoencoder and K-means++, which can increase the useful information by applying some image processing and automatically extract features by Multi-autoencoders. From the experimental results, it was clarified that the proposed method showed better clustering accuracy than the conventional methods with significant differences. In the future, we will make an application of the proposed method, for example, medical image analysis where the number of data with correct labels is limited.

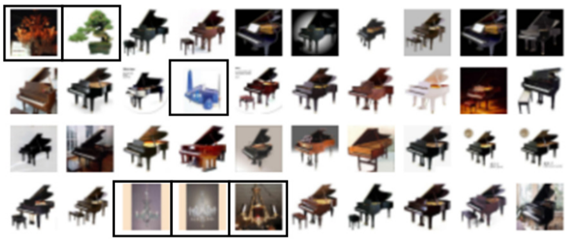

Fig. 4. Example of the generated cluster

References

1. Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature*. **521** (2015) 436–444.
2. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio and P. A. Manzagol, Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, *Journal of Machine Learning Research*, **11** (2010) 3371–3408.
3. C. Song, F. Liu, Y. Huang, L. Wang and T. Tan, Auto-encoder Based Data Clustering, *CIARP2013, Part I, LNCS 8258* (2013) pp. 117–124.
4. N. Kanopoulos, N. Vasanthavada and R. L. Baker, Design of an image edge detection filter using the Sobel operator. *IEEE Journal of solid-state circuits*, **23**(2) (1988) 358-367.
5. L. Fei-Fei, R. Fergus and P. Perona, One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, **28**(4) (2006) 594–611.
6. D. Arthur and Dergei Vassilvitskii, k-means++: The Advantages of Careful Seeding, *Proc. of the 18th annual ACM-SIAM symposium on Discrete algorithm* (2007), pp. 1027–1035
7. J. MacQueen, Some methods for classification and analysis of multivariate observations, *Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, (1967), pp. 281–297.
8. N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005*, **1** (2005) pp. 886–893.
9. D. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014)