International Journal on Emerging Research Areas (IJERA) *ISSN:2230-9993*

DOI:

CNN model to classify visually similar Images

Ms Tintu Alphonsa Thomas
Asst. Professor, Dept. of Computer Science & Engineering, Amal Jyothi College of Engineering, Kanjirappally, Kerala.

Ms. Anishamol Abraham
Asst. Professor, Dept. of Computer Science & Engineering, Amal Jyothi College of Engineering, Kanjirappally, Kerala.

Abstract

To cluster a large set of unlabelled images in the absence of training data remains a difficult task. A convolutional neural network (CNN) is suggested as a solution to clustering in order to deal with this issue. The suggested approach applies deep learning immediately to test data after receiving an input image set, as opposed to first building a training data set and then training a neural network on it.

Keywords-Convolutional neural network (CNN), Deep learning, Image clustering

I. INTRODUCTION

Image clustering is a fundamental problem for many image processing and computer vision applications. Nowadays, a huge number of images have been uploaded to cloud for sharing or storage. How to efficiently organize such large-scale image data is a challenging issue. In general, most research works on large-scale image clustering were based on feature encoding, such as hashing, which can largely reduce the dimensionality of image features so as to make large-scale clustering possible. However, reducing the dimensionality of features is equivalent to decreasing the representational power, leading to unsatisfactory clustering performance. Besides, the hash-based approaches usually assume that features are extracted before hash encoding. Different feature representations might lead to redesigning hash functions because of different number of dimensions or value ranges of feature vectors.

Clustering methods can be roughly categorized into hierarchical clustering and centroid-based clustering. The most popular algorithms for hierarchical clustering are agglomerative clustering. In agglomerative clustering, initially, individual samples in input data are considered as a cluster containing a single sample. Then, in each iteration, the two closest clusters in the raw or feature domain are merged into a cluster and the centroid of the merged cluster is computed accordingly. By iteratively merging the two closest clusters and updating the associated cluster centroids each time, we finally obtain the desired number of clusters and the corresponding centroids, which is, however, computationally very expensive for a large dataset. In contrast, centroid-based clustering (e.g., k-means and spectral clustering) randomly picks k samples from the input data as initial cluster centroids. Then, each sample finds its closest cluster centroid and is assigned with the corresponding cluster label. As a result, the cluster centroids are updated according to the clustering result. The clustering and centroid updating are sequentially iterated until converging to a solution point. To further improve clustering performance, some advanced techniques such as spectral clustering and matrix factorization were proposed to map visual features to another discriminative feature space to boost clustering performance. Such centroid-based clustering is more suitable for large-scale data clustering than hierarchical clustering due to less memory usage and computational power requirements. The effectiveness

DOI:

of centroid-based clustering, nevertheless, highly relies on feature representational power.

II. Convolutional neural networks

Convolutional neural networks (CNNs) have been widely used in image recognition tasks. There are several reasons why CNN models are gaining popularity. For instance, CNN models have a deeper architecture; thus, it can analyse more detailed features in images than traditional neural networks (NNs). The typical structure of CNNs has three layers, that is, the input layer, hidden layer, and output layer. Raw images are fed into the input layer and further transferred to the hidden layer to perform feature extraction tasks. The hidden layer is composed of convolution, pooling and fully connected layers. Feature extraction in the convolution and pooling layers is the main characteristic of a CNN model. The convolution layer can automatically extract features from raw images by convoluting the raw images with a sliding window of a fixed-size filter. The function of the pooling layer is to reduce the dimensionality of feature maps while retaining relevant features. Generally, the fully connected layer is the last layer of a CNN model, which then flattens feature maps to perform classification

III. Designing an Image Clustering Algorithm.

Consider the data set from Kaggle, for creating an image clustering model to classify the given images into two categories, namely toys or consumer products, and read the text written on the consumer products. The following are a few images from this data set.

For feature extraction, algorithm used Resnet50, pre-trained CNN model to remove the final layer of neurons used for prediction of classes. Image was put into the CNN and retrieved the feature vector as an output, which is a flattened array of all the feature maps learned by the CNN at the second last layer of Resnet50. This output vector can be given to a K means clustering algorithm which classify our images into the desired number of classes. The following figure shows the clusters that were made by this approach.

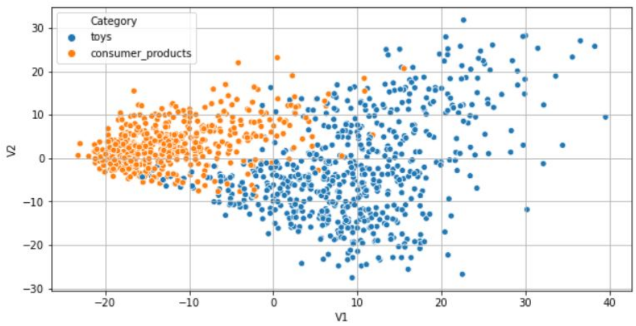

Figure 1: Clusters generated using K-Means.

DOI:

The code for this visualization is as follows

```
## lets make this a dataFrame
import seaborn as sb
import matplotlib.pyplot as plt
dimReducedDataFrame= pd.DataFrame
                        (Training_Feature_vector)
dimReducedDataFrame = dimReducedDataFrame.rename
                        (columns = { 0: 'V1', 1 : 'V2'})
dimReducedDataFrame['Category'] = list
                              (df['Class_of_image'])
plt.figure(figsize = (10, 5))
sb.scatterplot(data = dimReducedDataFrame,
                  x = 'V1', y = 'V2',hue = 'Category')
plt.grid(True)
plt.show()
```

IV. CONCLUSION

In conclusion, CNN models have shown promising results in classifying visually similar images. Different CNN architectures and techniques have been developed to achieve this task, and visualization techniques have been developed to aid in understanding the learned features of CNNs. These advances in CNNs will continue to contribute to the development of more accurate and efficient models for classifying visually similar images. This paper proposes an image clustering method to cluster visually similar images together using deep learning and clustering. It is entirely possible to cluster similar images together without even the need to create a data set and training a CNN on it.

References

[1] Chih-Chung Hsu, Chia-Wen Lin “CNN-Based Joint Clustering and Representation Learning with Feature Drift Compensation for Large-Scale Image Data”.

[2] Y. Zhang, Y. Zhang, J. Zhang, and Q. Dai, “CCR: Clustering and collaborative representation for fast single image super-resolution,” IEEE Trans. Multimedia, vol. 18, no. 3, pp. 405–417, Mar. 2016.

[3] Y. Gdalyahu, D. Weinshall, and M. Werman, “Self-organization in vision: stochastic clustering for image segmentation, perceptual grouping, and image database organization,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 10, pp. 1053–1074, Oct. 2001.

[4] K. C. Gowda and G. Krishna, “Agglomerative clustering using the concept of mutual nearest neighborhood,” Pattern Recog., vol. 10, pp. 105–112.

[5] Y. Gong et al., “Web scale photo hash clustering on a single machine,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Boston, MA, USA, Jun. 2015, pp. 19–27.

[6] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On spectral clustering: Analysis and an algorithm,” in Proc. Conf. Neural Inf. Process. Syst., V

[7] Che-Lun Hung, Deep learning in biomedical informatics, in Intelligent Nanotechnology, 2023.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton,"Learning Deep Representations for Visual Recognition" , 2012.

[9] Karen Simonyan and Andrew Zisserman ,"Very Deep Convolutional Networks for Large-Scale Image Recognition" ,2015.

[10] Christian Szegedy, Sergey Ioffe, and Vincent Vanhoucke, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning" ,2016.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun ,"Deep Residual Learning for Image Recognition" ,2016.

[12] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger ,"DenseNet: Densely Connected Convolutional Networks" ,2017.