Shapley value in convolutional neural networks (CNNs): A Comparative Study

Seyedamir Shobeiri

Department of Computer Science, Islamic Azad University of Zanjan, Zanjan, Iran.

seyedamir.shobeiri@iauz.ac.ir

Mojtaba Aajami

Department of Computer Science, Islamic Azad University of Zanjan, Zanjan, Iran.

aajami@iauz.ac.ir

***Abstract*—Deep learning models are in dire need of training data. This need can be addressed by encouraging data holders to contribute their data for training purpose. Data valuation is a mechanism that assigns a value reflecting a number to each data instances. The SHAP Value is a method for assigning payouts to players of coalition game depending on their contribution to the total payout that entails many criteria for the notion of data value. In this paper, the value of the SHAP parameter is calculated in different convolutional neural network for varieties of image datasets. Calculated SHAP value for each data instance shows whether data is high value or low value and it is different in each model. In other words, if you have an image in the VGG model and it is high value, necessarily, it is not high value in ResNet model. The results show that the value of data varies in each dataset and model.**

***Keywords*— Deep learning, SHAP Value.**

I. INTRODUCTION

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. The machine learning can be split into different classes of problems, Supervised learning and Unsupervised learning.In supervised learning, these algorithms, also called classifiers, receive data that has already been labeled. In fact, the data sent to the classification algorithms tells these algorithms which category each data should be placed in.

We do not tell the algorithm which category each data should fit into. In fact, we have no presumption about what category or group the existing data falls into, and the algorithm automatically detects the data classification. This is why these types of algorithms are called unsupervised learning algorithms.

The third type of algorithm, which can perhaps be classified as unsupervised algorithms, is a group called reinforced learning. In this type of algorithm, a machine (actually its controller program) is trained to make a specific decision, and the machine is based on its current position (set of available variables) and permissible actions (e.g. moving forward, moving Back and forth…) makes a decision that for the first time, this decision can be completely random and for each action or behavior that occurs, the system gives him a points from this feedback , The machine realizes whether it has made the right decision or not to repeat the same action the next time in that situation or try another action and behavior.

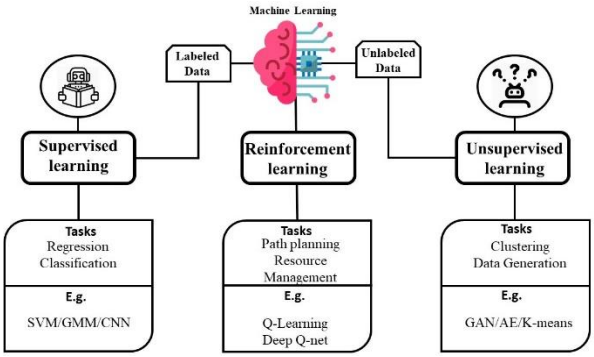

Fig. 1 Machine learning overview.

Deep Learning refers to a machine learning technique that establish artificial neural networks (ANNs) to imitate the structure and function of the human brain. Indeed, an artificial neural network, or neural network (NN), is a collection of connected computational units or nodes called neurons arranged in multiple computational layers that map a data input into a desired output. Each neuron applies a linear function to its inputs that sums up the products of weights and inputs. Subsequently, the output of this function is passed through an activation function. NN generates the desired output via feed-forward data flow and then updates the weights of each neuron by backpropagation of errors during the training phase.

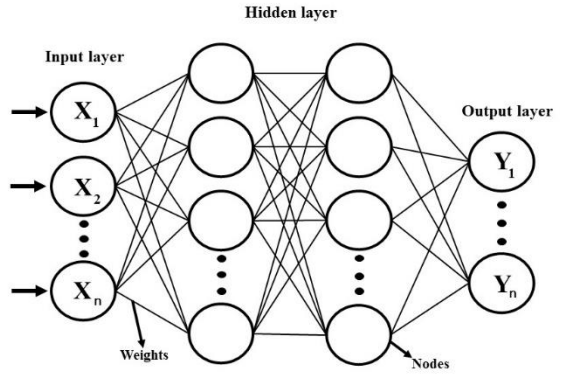

Fig. 2 Neural Network Overview.

In general, according to the employed NN architecture supervised deep learning can be categorized to the Recurrent neural networks (RNN) and Convolutional neural networks (CNN). RNN is a type of artificial neural network commonly used in speech recognition and natural language processing. Recurrent neural networks recognize data's sequential characteristics and use patterns to predict the next likely scenario. CNN is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data.

A Recurrent Neural Network is a type of neural network that contains loops, allowing information to be stored within the network. In short, Recurrent Neural Networks use their reasoning from previous experiences to inform the upcoming events. Recurrent models are valuable in their ability to sequence vectors, which opens up the API to performing more complicated tasks.

Convolutional neural network, is a deep learning neural network designed for processing structured arrays of data such as images. Convolutional neural networks are widely used in computer vision and have become the state of the art for many visual applications such as image classification, and have also found success in natural language processing for text classification.CNNs are very good at picking up on patterns in the input image, such as lines, gradients, circles, or even eyes and faces. It is this property that makes convolutional neural networks so powerful for computer vision. Unlike earlier computer vision algorithms, convolutional neural networks can operate directly on a raw image and do not need any preprocessing.A convolutional neural network is a feed-forward neural network, often with up to 20 or 30 layers. The power of a convolutional neural network comes from a special kind of layer called the convolutional layer.Convolutional neural networks contain many convolutional layers stacked on top of each other, each one capable of recognizing more sophisticated shapes. With three or four convolutional layers it is possible to recognize handwritten digits and with 25 layers it is possible to distinguish human faces.The usage of convolutional layers in a convolutional neural network mirrors the structure of the human visual cortex, where a series of layers process an incoming image and identify progressively more complex features.The architecture of a convolutional neural network is a multi-layered feed-forward neural network, made by stacking many hidden layers on top of each other in sequence. It is this sequential design that allows convolutional neural networks to learn hierarchical features. The hidden layers are typically convolutional layers followed by activation layers, some of them followed by pooling layers.A simple convolutional neural

network that aids understanding of the core design principles is the early convolutional neural network LeNet-5, published by Yann LeCun in 1998. LeNet is capable of recognizing handwritten characters.

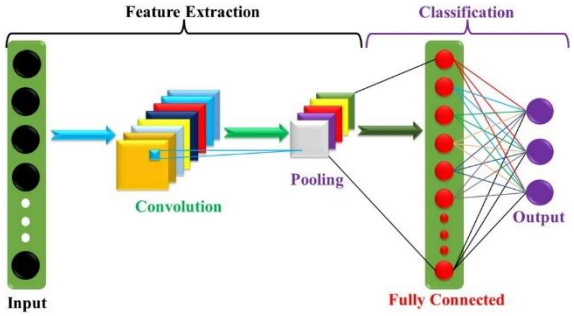

Fig. 3 Convolutional neural networks (CNN)Overview.

Machine learning models in general and deep learning in particular need to be trained with a wide variety of data in order to have acceptable performance. Hence there is a high demand from developers of machine learning models for training data. On the other hand, this data is owned by sectors that often do not voluntarily provide their data to model developers. In addition, we are interested in using data to training a model that will improve the efficiency of the model. Given this, there is a need for a mechanism that can evaluate the data needed and used to training the model. By using this method to provide incentives to data owners to share their data or to evaluate the data that is available for free before using it for training in terms of quality. In the seminal work, a mechanism for data valuation is provided. Inspired by Shapley Value method, this mechanism has presented a formula that can be used to numerically express the value of a specific data in improving the performance of a specific model. In this article, we have selected several famous CNN models that are widely used in various image recognition applications. we have calculated Shapley Value for two datasets, ImageNet and 10 Monkey Species, in each of these models. The rest of paper is organized as follows. Section II describes the architecture of several well-known CNNs.

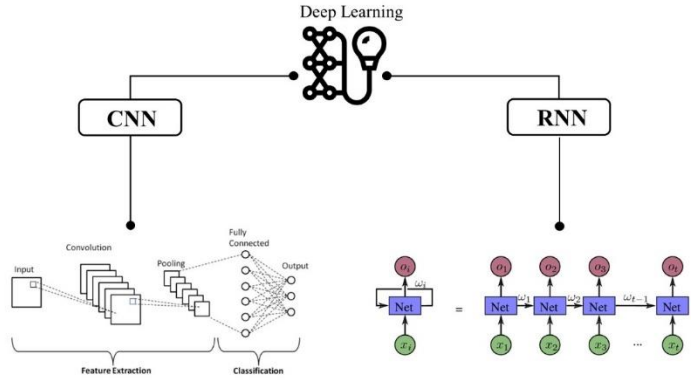

Fig. 4 NN architecture supervised deep learning Overview.

II. prolific CNN architectures

This paper seeks to empirically establish a correspondence between the shapely value and the components of a CNN based learning systems including model, training data set. For this purpose, we selected some CNNs as the representatives of the models that are widely used in practice.

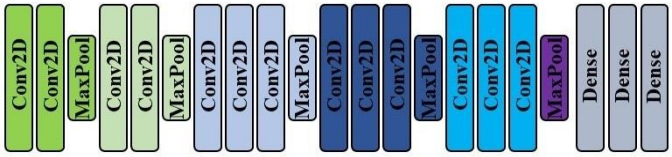

Fig. 5 vgg16 architecture.

Fixed size 224 x 224 RGB image is input of cov1 layer. The image through a convolutional.3*3 layers, which is the smallest size to capture the notion of left/right, up/down, center. it also utilizes 1×1 convolution filters, which as a linear transformation of the input channels. The convolution stride is fixed to 1 pixel; the spatial padding of convolution layer input is such that the spatial resolution is preserved after convolution, i.e., the padding is 1-pixel for 3×3 convolution layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the convolution. Max-pooling is performed over a 2×2-pixel window, with stride 2. Three Fully-Connected (FC) layers follow a stack of convolutional layers (which has a different depth in different architectures): the first two have 4096 channels each, the third performs 1000-way ILSVRC classification and thus contains 1000 channels .The final layer is

the soft-max layer. The configuration of the fully connected layers is the same in all networks.

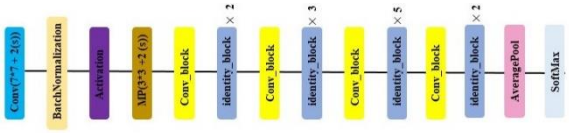

Fig. 6 ResNet50 architecture.

A convolution with a kernel size of 7 * 7 and 64 different kernels all with a stride of size 2 giving us 1 layer. Then we observe max pooling with a stride size of 2.In the next convolution there is a 1 * 1,64 kernel following this a 3 * 3,64 kernel and at last a 1 * 1,256 kernel, These three layers are repeated in total 3 time so giving us 9 layers in this step.Next step,we observe kernel of 1 * 1,128 after that a kernel of 3 * 3,128 and at last a kernel of 1 * 1,512 this step was repeated 4 time so giving us 12 layers in this step.so there is a kernal of 1 * 1,256 and 2 more kernels with 3 * 3,256 and 1 * 1,1024 and this is repeated 6 time giving us a total of 18 layers.again a 1 * 1,512 kernel with two more of 3 * 3,512 and 1 * 1,2048 and this was repeated 3 times giving us a total of 9 layers**.** Finally, we do an average pool and end it with a fully connected layer consist of 1000 nodes and at the end a Softmax function so this gives us 1 layer.Actually, we ignore activation functions and the max/ average pooling layers.as a result, it gives us a 1 + 9 + 12 + 18 + 9 + 1 = 50 layers Deep CNN.

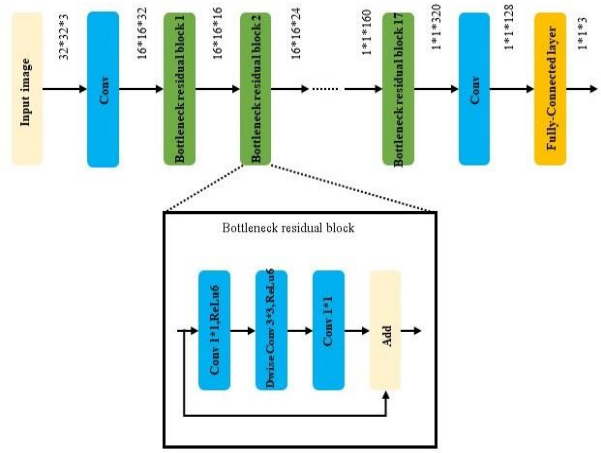

Fig. 7 MobileNetV2 architecture.

MobileNet-V2 is one of the efficient models from the MobileNet group. Two types of blocks are in the MobileNet-V2.one of them is residual block with stride of 1. Other one is block with stride of 2 for downsizing. There are three layers for two types of blocks.1*1 Convolution with ReLU6 is the first layer. Depthwise convolution is second layer. 1*1 convolution but without any non-linearity is third layer**.** *t is* expansion factor and *t*=6 for all main experiments. internal output would get 64×*t*=64×6=384 channels if the input got 64 channels.

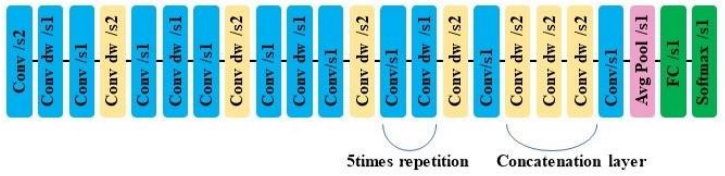

Fig. 8 MobileNet architecture.

MobileNet is a simple architecture. it uses depthwise divisible convolutions to build lightweight deep CNNs and provides a model for embedded vision applications and mobile.As shown in Figure 8 ,MobileNet structure is depthwise divisible filters. Depthwise divisible convolution filters are composed of depthwise convolution filters and point convolution filters. The depthwise convolution filter performs a single convolution on each input channel, and the point convolution filter combines the output of depthwise convolution linearly with $1 * 1$ convolutions.

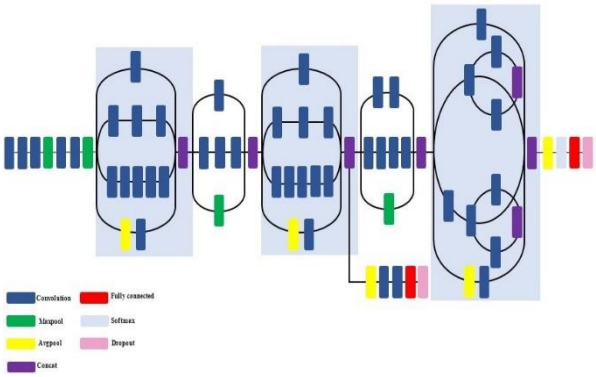

Fig. 9 InceptionV3 architecture.

Inception-v3 is a CNN with 48 layers for help in image, and start as a module of GoogleNet. The third edition of Google's Inception CNN is

Inception-V3, at first, it introduced for the ImageNet Recognition Challenge. Inception assists classification of objects in the computer vision.

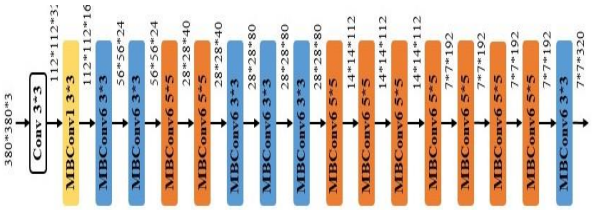

Fig. 10 EfficientNet B0 architecture.

EfficientNet-b0 is a CNN that is trained on more than a million images from the ImageNet database. The model can classify images into 1000 object categories, such as camera, boat, bird. So, the network learned rich features. The input of network is 224*224.EfficientNet-B0 the developed by AutoML MNAS.

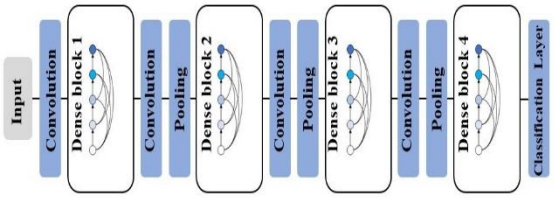

Fig. 11 DenseNet169 architecture.

III. Shapley Value

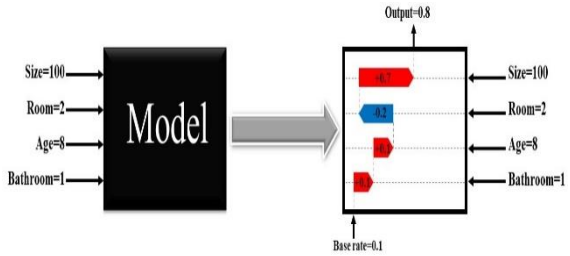

Fig. 12 Shapley Value example.

Shap Value performs the equitable data valuation in supervised machine learning. For a given set of training data points D and a performance metric, the "Shap Value" value $\emptyset_i$ of a data point $x_i \in D$ is defined as:

$$\emptyset_i = \sum_{S \subseteq D\{x_i\}} \frac{V(S \cup \{x_i\}) - V(S)}{\binom{|D|-1}{|S|}}$$

Where V(S)is the performance of the model trained on subset S of the data. V(S)is the prediction accuracy on the validation set. Intuitively, the Shapley value of a data point is a weighted average of its marginal contribution to subsets of the rest of the dataset. As a result, it can be used as a measure of data quality: a data point with a high Shapley value is one that improves the model's performance if we add it to most subsets of the data, while a data point with a negative value on average hurts the performance of the model. Exact computation of Eq. requires an exponential number of computations in the size of the dataset, which is infeasible in most realistic settings. In fact, High value indicate high quality of image and correct label while low value represents low quality of image and incorrect label.

Finally, SHAP Value has three outputs, which are value, growth rate, and main data, respectively. Value: An array that each cell represents a pixel, each cell of the array contains another array that contains three cells and represents the RGB effect as shown: Value = [[R, G, B], [R, G, B], [R, G, B], …] and the growth rate, which is a base number, and our main data, which is the original values of our image. How to calculate the value of an image is as follows:

$$Value = (\sum_{i=0}^{n}([R_i + G_i + B_i])) + \text{Base}$$

According to the above formula i = 0 because the array cells start from zero and N is the number of pixels, R represents the effect of red, G represents the effect of green, B represents the effect of blue, which indicates each of these pixel colors. How effective the image has been

in our model is that the sum of the effects of red, green, and blue with the base, which represents the growth rate, reflects the value of image. next, two datasets are tested that whether the value of an image is always the same or depends on another factor.

IV. Evolution and results

Due to the large volume of images, we randomly selected 50 images from each dataset and created two datasets smaller than ImageNet and 10Moneky Species. Images in the datasets are numbered from 0 to 49, and to refer to each image, Refer to the relevant number. In this article, we use the first formula, we obtained the value of each color of each pixel (RGB), and then use the second formula, we obtained the value of each image on different Convolutional neural networks, include VGG16, ResNet50, DenseNet169, MobileNet, EfficientNetB0, MobileNetV2. The table below has two rows that represent high quality and low-quality images and six columns that represent different architectures. The numbers you see in the tables indicate the number of images in each dataset. The numbers in the first row indicate the high-quality data of datasets in the architecture and the second row indicates low quality data in the same architecture.

Table 1: ImageNet Datasets

	VGG16	ResNet50	DenseNet169	MobileNet	EfficientNetB0	MobileNetV2
High Value	3	3	3	3	11	31
Low Value	34	25	18	32	25	18

Table 2: 10 Monkey Species Datasets

	VGG16	ResNet50	DenseNet169	MobileNet	EfficientNetB0	MobileNetV2
High Value	34	30	19	25	32	31
Low Value	46	43	36	38	46	37

V. Conclusions

Training data is used for training deep learning models. This need can be addressed by encouraging data holders to contribute their data for training purpose. We need a method for numerical evaluation of data that the Data valuation method is very efficient. The Shap Value is a way to evaluate each data relative to its contribution to deep learning models. According to the above experiments, we realized that the value of each image is very different from the network architecture and the value of each image is directly related to the architecture of that network.

References

Ghorbani, A., & Zou, J. (2019, May). Data shapley: Equitable valuation of data for machine learning. In *International Conference on Machine Learning* (pp. 2242-2251). PMLR.

Lundberg, S. (2018). *SHAP documentation*. SHAP. https://shap.readthedocs.io/en/latest/index.html