Journal of Engineering and Applied Sciences 14 (7): 2272-2275, 2019
ISSN: 1816-949X

The Definition of Convolution in Deep Learning by using Matrix

Hwajoon Kim
Department of General Education, Kyungdong University, 482-010 Yangju, South Korea

Abstract: We would like to give a mathematical concept in the establishment of definition of convolution in deep learning and the mathematical tool used is a matrix. The methodology used is to analyze the intersection of the mathematical meaning and the engineering meaning of the concept of convolution and through it we would like to establish the definition of convolution to give practical help to CNN technology of deep learning.

Key words: Convolution, deep learning, neural networkm, tool, matrix, practical

INTRODUCTION

We would like to contribute to the development of deep learning technology by accurately establishing the concept of convolution in the Convolutional Neural Network (CNN) a core technology of deep learning. Through the methodology of the research that compares the concepts of convolution in mathematics, electrical electronics and Artificial Intelligence (AI) we try to establish the definition of the most reasonable form of convolution in deep learning and study its nature.

The core of this CNN technique is the convolution which applies the weight to the receptive fields only and it transforms the original data into a feature map. This process is called convolution. However, since, the precise concept of convolution in CNN has not yet been established we intend to establish the concept of convolution in deep learning through this research. If this concept is established there are many research topics that will contribute to the development of CNN I would like to mention the hint in this study. Those studying mathematics are always approaching mathematically and those studying engineering always have an engineering approach. The intersection is quite difficult to find in these approaches. In this study we analyze the intersection of the mathematical meaning and the engineering meaning of the concept of convolution and establishes the concept to give practical help to CNN technology of deep learning.

On the other hand, the deep learning is a theory that was not developed at the times because the basic research was done 25 years ago and there was no mathematical theoretical system at that times and there was no computer system to handle the computation. As a part of representation learning, deep learning has started from attracting people's attention with victory in the Baduk match of Alphaho and Sd. Lee and there have been many studies to deal with overfitting.

The definition of convolution in deep learning by using matrix: This study requires a basic knowledge of the Convolutional Neural Network (CNN), so, let us introduce it first. Deep learning (Goodfellow *et al*., 2016; Graupe, 2016, LeCun and Ranzato, 2013; LeCun *et al*., 2015; Schmidhuber, 2015; Silver *et al*., 2016; Weston *et al*., 2012; Wiley, 2016) means deep neural network where one or more hidden layers are expressed as deep in the Feed-Forward Neural Network (FFNN). FFNN compute: $y = WX+b$ basically where X is input and W is weight. And after FFNN applies the activation function here. This process is called a layer. After stacking several layers and finally stacking the output layers it is called Multi-Layer Perceptron (MLP). This MLP performs this process repeatedly to find the optimal space to solve a certain problem. This process is called the representation learning. This CNN applies weight to a portion of what we call receptive fields whereas the Neural Network (NN) applies weights to all nodes. This has a form similar to what the big data problem can say as "what will we throw away?".

For example when processing a 28*28 matrix it sweeps to a 3*3 kernel matrix to transform the data into a different form. This process is called the convolution. This convolution is a multiplication of the elements corresponding to the window and the concept of convolution is different from it in mathematics. However, it is the researchers view that it can be interpreted in essentially the same meaning and providing the theoretical framework by this study is the main research goal of this study. If you sweep the image to the kernel window we get an image other than the original one. This distorted image is called a feature map. Note that the pixels are represented by a matrix and each pixel consists of numbers from 0-255 to represent all colors.

In a word in CNN, a convolution can be expressed as a tool that creates a feature map from the original image. When CNN has input it creates several feature maps by using the kernel matrix and then proceeds with classification. The kernel matrix w can be learnable in WX just as the weight W was learnable in WX+b of MLP's where X is an image. This means that CNN learns itself by using back propagation a process of calculating errors and then sequentially updating them. The last step of CNN is the pooling. This is the research of reducing size which is to transform the image of the convolution into a small image by extracting a representative from each part. For example in short if we have max-pooling a matrix:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

we obtain a matrix (4) = 4 of 1 row and 1 column (just a simple task of drawing the largest number). This pooling is simple but it makes a contribution to enhance performance. Because it is the principle that the resolution is increased when the screen is reduced. Silver *et al.* (2016) has shown that features are invariant in fine translation and reduce noise. The currently used pooling method is max-pooling which is quite simple as the above but has a reasonable effect. There are good research ideas in here. Changing the pooling method will change the resolution. For classification you can get a high resolution image by improving the pooling method. It is a simple research topic but it is not, so easy to find a new pooling method which is improved. This is the research topic to be confirmed by the mathematical method and simulation. Besides this there are numbers of factors that relate to performance contributions, e.g., the number of total layers, the number of receptive fields, the size of receptive fields, the size of pooling fields and the size of feature maps and number of them. Since, the concept of convolution has a lot of gap between mathematics and information technology we would like to connect these two concepts naturally. The definition of convolution in mathematics.

MATERIALS AND METHODS

Definition 1: The convolution of f and g is defined by the integral:

$$\left(f^*g(t)\right) = \int_0^t f(\tau)g(t-\tau)d\tau$$

In order to connect to IT, we would like to limit the integration interval to (0t) and have to discretion.

The interpretation of convolution in electrical engineering: Convolution is an important technique in signal and image processing. It provides a means of calculating the response or output of a system to an arbitrary input signal if the impulse response is known (Croft *et al.*, 2001).

As mentioned above, Croft *et al.* (2001) describes the role of the convolution well and its definition is interpreted as overlapping parts by drawing the given function f and g. This means that the multiplication in convolution can be interpreted as the concept of intersection. Consequently, the mathematical definition and its meaning in engineering are not contradictory and the result of the calculation is always the same. Generally, the following expression always holds:

$$F(f^*g) = F(f)F(g)$$

for f(t) = F(W) where F is a Fourier transform. The above equation is also true for the laplace transform (Kreyszig, 2011) and a generalized integral transform (Kim, 2017) as well. That is:

$$\pounds(f^*g) = \pounds(f)\pounds(g)$$

In engineering, the convolution in the time domain corresponds to the multiplication in the frequency domain. That is the way to obtain convolution is to use a graph and the calculation is performed when there are overlapping parts (calculation of the enclosed area by integral) and when there is no overlapping part, the value of convolution is interpreted as 0. The result is equivalent to the value obtained by the mathematical formula. For example for u(t) is the unit step function, given:

$$f(t) = u(t)e^{-t},\ g(t) = u(t)e^{-2t}$$

The convolution is obtained as the following; First we draw a graph of f (τ) then draw a graph of g ($-\tau$). Next we translate it by t and calculate the value of the overlapped parts. The result is equal to the convolution calculated as:

$$\left(f^*g(t) = u(t)\left(e^{-t}-e^{-2t}\right)\right)$$

To apply this convolution to IT you have to discretion as follows:

$$f^*g = \sum_{\tau=0}^{t} f(\tau)g(t-\tau)$$

RESULTS AND DISCUSSION

The analysis of convolution in deep learning (for classification): Since, the definition of convolution in CNN has not yet been clarified we would like to proceed this study. Although, the definition of convolution is not clear, the way to obtain the currently used convolution is as follows: sweeping the original image with a matrix of 3*3 kernel matrices, transforms original data into a different shape. The concrete way to do this is to multiply each 3*3 part of the original matrix by the kernel matrix and add all components. This is the way in which we get the convolution. For example, when the original image is:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

Element-wise multiplying the kernel matrix:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

On the first 3*3 matrix:

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

We obtain the matrix:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Now, adding all of its components, we obtain 4. Next, element-wise multiplying the kernel matrix on the next 3*3 matrix:

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

and adds every components, we get 3. If we continue this process to the final matrix:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

we get 4. Consequently, the original matrix changes to:

$$\begin{pmatrix} 4 & 3 & 4 \\ 2 & 4 & 3 \\ 2 & 3 & 4 \end{pmatrix}$$

by using the convolution kernel. This is called the convolved feature map (the image obtained as a result of convolution). As mentioned before in the classification field of artificial intelligence, the pixel is treated as a matrix. Thus, there is a need to make the definition of the convolution to be matched with that of the mathematical convolution. Hence, we would like to introduce the definition of convolution by matrix.

Definition 2: The convolution of matrices A and B can be defined as the sum of all components of:

$$AB^T\chi_D(i, j)$$

where, T is the transpose, χ is characteristic function and:

$$D = \{(i, j)|i = j\}$$

Let us take a closer look at the above definition. To apply the convolution to IT we have to discretion as follows:

$$f*g = \sum_{\tau=0}^{t} f(\tau)g(t-\tau)$$

We note that the method of calculating convolution in engineering is the follows; Drawing a graph, the calculation is performed when there is overlapping part. If there is no overlapping part, the value of convolution is interpreted as 0. In other words we plot the graph of $f(\tau)$, draw the graph of $g(-\tau)$ and then translate it by t to calculate the overlapping value. Of course, this result is entirely consistent with the mathematical definition.

If the matrix representing a function f is A and the matrix representing a function g is B then the convolution of the functions f and g can be denoted by $AB^T\chi_D(i, j)$. Intuitively, the diagonal part of B^T corresponds to a graph of $g(t-\tau)$. The overlapping part of the graph can be interpreted as the concept of intersectio, that is the concept of multiplication. A little further as in the example of p6 in CNN of deep learning, the convolution is a tool to obtain the feature map from the original image data. The calculation method is to multiply by element-wise and add all of components. In formula if:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

and

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

Then, the convolution f and g denoted by f*g and defined by:

$$a_{11}b_{11}+a_{12}b_{12}+a_{13}b_{13}+, ..., +a_{32}b_{32}+a_{33}b_{33}$$

Of course, this is the same result with the sum of all components of:

$$AB^T\chi_D(i, j)$$

where, D = {(I, j)|i = j}. Here, note that functions can be interpreted as image.

CONCLUSION

The core technology of this deep learning is the Convolutional Neural Network (CNN). The most important concept among them is the convolution. In this study, we would like to establish the definition and concept of convolution by using matrix a rectangular array of numbers or letters.

REFERENCES

Croft, A., R. Davison and M. Hargreaves, 2001. Engineering Mathematics. Pearson Education Ltd, ýLondon, England, UK.,.

Goodfellow, I., Y. Bengio and A. Courville, 2016. Deep Learning. MIT Press, Massachusetts, United States, ISBN:9780262337373, Pages: 800.

Graupe, D., 2016. Deep Learning Neural Networks: Design and Case Studies. World Scientic, Singapore, ISBN:9789813146440, Pages: 263.

Kim, H., 2017. The intrinsic structure and properties of laplace- typed integral transforms. Math. Prob. Eng., 2017: 1-8.

Kreyszig, E., 2011. Advanced Engineering Mathematics. 10th Edn., John Willy & Sons, Hoboken, New Jersey, ISBN:9781118165096, Pages: 924.

LeCun, Y. and M. Ranzato, 2013. Deep learning tutorial. Proceedings of the International Conference on Machine Learning (ICML'13), June 16-21, 2013, Atlanta Marriott Marquis, Altanta, Georgia, pp: 1-29.

LeCun, Y., Y. Bengio and G. Hinton, 2015. Deep learning. Nature, 521: 436-444.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks, 61: 85-117.

Silver, D., A. Huang, C.J. Maddison, A. Guez and L. Sifre *et al.*, 2016. Mastering the game of Go with deep neural networks and tree search. Nat., 529: 484-489.

Weston, J., F. Ratle, H. Mobahi and R. Collobert, 2012. Deep Learning via Semi-Supervised Embedding. In: Neural Networks: Tricks of the Tradem, Montavon, G., G. Orr and K.R. Muller (Eds.). Springer, USA., pp: 639-655.

Wiley, J.F., 2016. R Deep Learning Essentials. Packt Publishing Ltd., Birmingham, England, UK., ISBN:978-1-78528-058-0, Pages: 153.