Optik - International Journal for Light and Electron Optics 172 (2018) 359–367

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.com/locate/ijleo

Original research article

Optimal hyperparameter tuning of convolutional neural networks based on the parameter-setting-free harmony search algorithm

Woo-Young Lee[a,b], Seung-Min Park[a], Kwee-Bo Sim[a,*]

[a] School of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Republic of Korea
[b] Intellivix Co., Ltd, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
CNN
Hyperparameter tuning
PSF-HS algorithm

ABSTRACT

Hyperparameters determine layer architecture in the feature extraction step of a convolutional neural network (CNN), and this affects classification accuracy and learning time. In this paper, we propose a method to improve CNN performance by hyperparameter tuning in the feature extraction step of CNN. In the proposed method, the hyperparameter is adjusted using a parameter-setting-free harmony search (PSF-HS) algorithm, which is a metaheuristic optimization method. In the PSF-HS algorithm, the hyperparameter to be adjusted is set as the harmony, and harmony memory is generated after generating the harmony. Harmony memory is updated based on the loss of a CNN. A simulation using CNN architecture with reference to LeNet-5 and a MNIST dataset, and a simulation using the CNN architecture with reference to CifarNet and a Cifar-10 dataset are performed. By two simulations, it is possible to improve the performance by tuning the hyperparameters in CNN architectures proposed in the past.

1. Introduction

At the world economic forum (WEF) held in Davos at January 2016, there the impact of the fourth industrial revolution (Industry 4.0) on economy and society was discussed. At the WEF held in January 2017, there was a debate about the realization of the fourth industrial revolution after 2016. The fourth industrial revolution is the next-generation industrial revolution led by artificial intelligence (AI), robot technology, and life sciences. The key concept is to converge technologies based on the Internet. The core technologies of the fourth industrial revolution include the Internet of Things (IoT) and AI [1].

AI, the core technology of the fourth industrial revolution, was brought in the 1950s with the question, “Can machines think?” by Alan Mathison Turing, a British mathematician and logician. Turing insisted that if a machine could respond like a real human, it would certainly be considered intelligent. AI implies that a computer can imitate intelligence by demonstrating abilities such as human perception, reasoning, learning, natural language understanding, and communication under various situations [2].

An artificial neural network (ANN) imitates the neuron connection structure in the human brain. It is used in functional reasoning, pattern recognition, and clustering. An ANN consists of an input layer, hidden layer, and output layer, and training is performed by adjusting the weight of each neuron [3].

Training an ANN requires a sufficient amount of training data and requires high-performance hardware. In the past, there was a lack of data and the hardware performance was unsatisfactory. By the early 2000s, the dark period of the ANN lasted; however with the recent development of graphics processing units (GPUs) optimized for big data and parallel operation, the neural network

* Corresponding author.
E-mail address: kbsim@cau.ac.kr (K.-B. Sim).
URL: http://alife.cau.ac.kr (K.-B. Sim).

https://doi.org/10.1016/j.ijleo.2018.07.044
Received 12 May 2018; Accepted 13 July 2018
0030-4026/

algorithm has become popular.

Deep learning (DL) is a type of neural network developed from ANN. It is characterized by a deep depth of hidden layers. Examples of DL include the convolutional neural network (CNN), which has excellent performance in computer vision, and the recurrent neural network (RNN), which performs well in natural language processing.

CNNs proposed in the 1990s were machine learning models based on supervised learning. These CNNs consisted of an input layer, hidden layer, and output layer, similar to the ANN. The CNN has advantages in that it can directly operate on raw data without extracting the characteristics of the data. The CNN achieves because it includes the step of extracting features in the inside. In constructing the CNN architecture, hyperparameters such as the size of the convolution kernel, number of channels, padding, and stride influence the performance of the CNN because they determine the layer structure, including the size of the feature map. To obtain the expected result using CNN, an appropriate hyperparameter should be used. However, at present there is no clear rule for optimizing the hyperparameters of a CNN, and hyperparameters are mostly determined by the experience and intuition of the designer. In fact, in the structure of the proposed CNN, the size of all convolution kernels is set to 5 in the LeNet-5 architecture, while in Alexnet, the size of the convolution kernels is set to 11, 5, and 3. In the ZFNet architecture, the size of the convolution kernel at 7, 5, and 3, and the hyperparameters in each layer are determined according to the designer's intuition [4–6].

In this paper, we propose a method to tune the hyperparameters of the feature extraction step of a CNN using a parameter-setting-free harmony search (PSF-HS) algorithm. In the PSF-HS algorithm, a variable to be optimized is called a harmony. We hypothesized that if we set the hyperparameters to harmonies and apply the PSF-HS algorithm to update the harmonies, we can obtain optimal hyperparameters.

2. Related works

2.1. Convolutional neural network

The CNN is a type of ANN, and it is used mainly in the field of computer vision. The CNN classifies and recognizes objects and texts in images because it shows excellent performance for feature extraction, classification, and pattern recognition of input data. A CNN is composed of an input layer, a hidden layer, and an output layer. The hidden layer of the CNN consists of extracting features of input data and classifying extracted features. The convolutional layer of the hidden layer of the CNN extracts features of input data through a convolution operation. In this layer, a convolution filter with a weight is used to extract features by applying a convolution operation to the input data. The pooling layer is the layer that abstracts the input data. The pooling types are max pooling and average pooling. Max pooling is used for image recognition. The pooling layer is characterized by having no parameters to train, and no change in the number of channels. The fully connected layer classifies the features extracted in the feature extraction step by a general multilayer neural network architecture [7–9].

Since a general ANN computes the output for the raw input data without considering the features of the data, there are many neurons to be trained. When classifying similar data, it is possible to classify the image into a completely different image if the color information of the image and the location information are different. On the other hand, the CNN computes the output for the features of the input data. Since size of the features of the input data are smaller than those of the input data, the number of neurons to be trained is smaller than that of the general ANN [10,11].

Training the CNN involves adjusting the weights of the convolution kernel and the weights of the neurons in the fully connected layer. A CNN generally uses algorithms such as error backpropagation and gradient descent to adjust the weights of the convolution kernel and the neurons in the fully connected layer [12].

2.2. Parameter-setting-free harmony search algorithm

The harmony search (HS) algorithm is a metaheuristic algorithm that imitates a method to find optimal chords for improvising music and obtains an optimal solution to a problem. The HS algorithm uses a space that stores a harmony called harmony memory (HM). The HM stores the harmonies that are judged to be good, and updates the method by excluding bad harmonies to find the optimal solution [13,14].

In the HS algorithm, the solution to be searched is set as a harmony, and variables are generated by three methods: random selection, harmony memory consideration, and pitch adjustment. The random selection method randomly generates a harmony, and harmony memory consideration uses the harmonies of the HM. The pitch adjustment generates new harmonies by modifying the harmonies of the HM [15].

The probability of generating harmonies using the harmony memory consideration method (harmony memory considering Rate, HMCR) and the pitch adjustment method (pitch adjusting rate, PAR) affects the generation of the harmony. Thus, HMCR and PAR affect the performance of the HS algorithm. In the general HS algorithm, HMCR and PAR are initially set, and a harmony is generated by selecting either the random selection, harmony memory consideration, or pitch adjustment methods in the HS algorithm according to the set probability.

The generated harmonies are compared with the worst-case harmonies in the HM. If their performance is better, the newly created harmonies are included in the HM, and the worst-case harmonies in the existing harmony memory are excluded. In the opposite case, the HS algorithm does not update the HM and does not use the newly created harmonies. The HS algorithm is known to have excellent local convergence and early convergence stabilities. The HS algorithm is shown in Algorithm 1.