An Enhanced Neural Network Approach for Numeral Recognition

Anita Venugopal[1], Ashraf Ali[2]
[1]Dhofar University, Salalah, Postal Code 211, Oman

[2]Faculty of Computer Studies, Arab Open University, Kingdom of Bahrain

Abstract
Object classification is one of the main fields in neural networks and has attracted the interest of many researchers. Although there have been vast advancements in this area, still there are many challenges that are faced even in the current era due to its inefficiency in handling large data, linguistic and dimensional complexities. Powerful hardware and software approaches in Neural Networks such as Deep Neural Networks present efficient mechanisms and contribute a lot to the field of object recognition as well as to handle time series classification. Due to the high rate of accuracy in terms of prediction rate, a neural network is often preferred in applications that require identification, segmentation, and detection based on features. Neural networks self-learning ability has revolutionized computing power and has its application in numerous fields such as powering unmanned self-driving vehicles, speech recognition, etc. In this paper, the experiment is conducted to implement a neural approach to identify numbers in different formats without human intervention. Measures are taken to improve the efficiency of the machines to classify and identify numbers. Experimental results show the importance of having training sets to achieve better recognition accuracy.

Key words:
OCR, Classification, Neural Network, Learning Rate, Recognition

1. Introduction

Human brains are so astonishing that they can effortlessly recognize any characters with ease. The neurons in our brains are interconnected to billions of other neurons and are well adapted to reading any complex characters in a fraction of a second. Though difficult, visual recognition for machines is made possible through the computer programming neural network approach. Dataset is collected, the network is designed and trained with the dataset samples. The network is trained with a greater number of training samples to improve the prediction accuracy.

Similar to human neurons, the mathematical model of artificial neural network technology uses perceptron's which takes several inputs and produces a single output, a 0 or 1. Weights are added and assumed threshold values are adjusted till the output is less than 0 or greater than 1. Artificial perceptron's are programmed for decision-making by adding weights. Weights and thresholds are adjusted to depict the decision-making ability of humans. Different layers of perceptron's are designed for decision-making. Initially, weights are added and the perceptron's in the first layer outputs simple decisions. The perceptron's in the subsequent layers are responsible for making more and more complex decisions. Learning algorithms adjust bias or weight values to produce output which is the same as the input. In this paper, a perceptron with a sigmoid neuron is used to improve the prediction rate. The advantage of a sigmoid neuron over a perceptron is that a perceptron takes an input and produces either 0 or 1 output, whereas, a sigmoid neuron outputs all the possible probabilities between 0 and 1[4].

Neural Network (NN) architecture consists of nodes arranged in layers where a group of learning algorithms is used as opposed to a single algorithm. It is used to learn complex predictions consisting of multi-layer NN with several hidden units [5]. NN techniques have successfully solved several application problems related to the classification of handwritten numerals of the MNIST data set with an error rate of 0.21% [6]. Application areas also include recognition of an image, speech [7] [8], language [9], computational biology [10], pattern recognition, clustering, time series prediction, function approximation, etc.

In this paper, section II presents a literature survey, section III presents the preliminaries and existent neural network approaches for character recognition. Section IV discusses the proposed neural approach used for designing the network. Section V introduces network implementation and classification techniques. Experiment results are shown in section VI and section VII marks the conclusion.

2. Literature Review

Artificial neural network models have been popular since the 1950s [11] however, the latest advancement of deep learning has its common characteristics of SLNN (supervised-learning neural network) and ULNN (unsupervised-learning neural network) approaches with several hidden layer neurons using a combination of

Manuscript received March 5, 2022
Manuscript revised March 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.3.9

regulated Boltzmann system, backpropagation and error gradients [12]. The first model of a neuron was shaped in 1943 [13]. A logical calculus approach inherent in nervous activity was proposed by Bull [14]. However, this model does not mimic the biological neuron and did not perform learning. The first biologically motivated artificial neuron with unsupervised learning abilities was introduced by Hebb in 1949[15]. The real notion of perceptron was developed in 1957 by Rosenblatt [16]. Perceptron's are single-layer neural networks with the ability to function as a classifier.

Neural networks applications are numerous and varied from financial markets to economic growth analysis [17]. Hyup Roh in 2007 introduced combination models with neural networks time series for predictions of the fluctuations in the stock price. The results of the experiments by Veri and Baba in 2013 declare that the most appropriate design for network design consists of 40 training datasets resulting in 95% accuracy of data. NN with its powerful algorithmic functions finds its application in fields like pattern recognition, memory, mapping, etc. [18].

Studies show that template matching is the simplest approach used for the recognition of characters but is sensitive to noise/illumination. Research studies show that diagonal-based feature extraction is ideal for handwritten NN classification [19]. Sharma, Om Prakash et al proposed an enhanced feature extraction model utilizing Euler number for alphabet/pattern recognition [20]. Research studies by Sahu have proposed two main classes of features namely, statistical and structural approaches for feature extraction [21].

The immense success rate of the recurrent network model on handwriting recognition, speech [22] [23], and image recognition [24] has inspired this study. Since recurrent models lack feature extraction ability, a hybrid architecture such as a combination of the recurrent and convolutional network model is considered ideal in the recognition of characters [25].

3. Neural Network Approach

Neural networks/deep learning can tackle problems associated with object detection/recognition, speech recognition, and other language/morphological processing. In the traditional programming approach, programs are written and fed to the computer to solve a problem. Whereas, in a neural network, machines are trained to acquire the ability to classify and recognize objects on their own to the problem at hand. Mathematical algorithms are used to train computers so that machines develop the ability to make decisions on their own. FFN, CNN, and RNN are a few of the popular neural networks that form the fundamental trained models in deep learning.

3.1 Artificial Neural Network (ANN)

ANN or Feed-Forward Neural network (FFN) is called so because the flow of information is always forward. There are no loops. In the FFN model, the output from one layer is passed on to the subsequent layers. The basic model of artificial NN comprises three layers namely input layer, hidden layer, and output layer. Input is fed through the input layer, the hidden layer is responsible for all the processing and decision making like classification and recognition, and the third layer named output displays the recognized result. Each layer learns from the data set and certain weights are added. The artificial neural network can handle nonlinear functions. This is mainly possible because of the activation function which can introduce nonlinear properties to the network and hence ANN acquires the ability to learn any complex tasks/relationship between the first and last layer [1].

3.2 Convolution Neural Network (CNN)

CNN model otherwise is also called deep learning is distributed image representation. To identify x different input images, CNN extracts y different features from the input image. It can extract the spatial features from the input. This helps in identifying the object, its location, its relation with other objects in the image precisely [2].

3.3 Recurrent Neural Network (RNN)

RNN is derived from FFN, feedback or loops are possible. Learning algorithms of RNN make it useful in the field of handwritten characters or speech recognition.

To obtain better results, hybrid methods are used to design systems. The output generated by CNN is continued by a recurrent neural network (RNN) in this system. Features extracted by CNN are input to RNN. RNN can process the data sequentially. RNN has short-time memory and hence cannot retain information [26]. To solve this vanishing gradient issue, long-term short-memory (LSTM) architecture is adopted. It is designed to retain information for long periods. LSTM can backpropagate through layers. LSTM exhibits superior performance and can deal with variations in different types of lines using its two-dimensional architecture. This architecture blends one-directional networks. Each network forwards the output to its successor.

4. Network Design

Handwritten digits are collected for training and the network is developed. Input is gathered in grayscale format. The input image is binarized to its binary form. Using certain threshold values the image is divided into the foreground and background forms. The output image is either in black or white format. To minimize the processing duration, non-significant objects from the image are eliminated and only significant data is gathered using an integral approach. The resultant image is free from noise and other disturbances.

The skewness due to scanning or writing style is corrected by finding the skew angles (horizontal or vertical slant in lines). The skew angle of deviation(θ) is calculated for up to ±700 to eliminate the maximum slant in the input. In this experiment, Hough transform is used to calculate the most likely horizontal and vertical (small lines) skewness.

To reduce the processing time, thinning is performed to remove extra pixels and skeletonize the text to one-pixel width. Input characters are separated into individual characters so that they can be identified.

4.1 Segmentation

Characters are separated and the segmented characters are classified and identified into individual digits. For machines, segmentation is a challenging task. The segmentation method is used to find out each trial segment. Horizontal and vertical projection technique is used for segmentation. Classifier with mark high score if all the segments are separated and low score is marked in the case of failure in detecting the segments. Fig. 1 shows the segmentation.

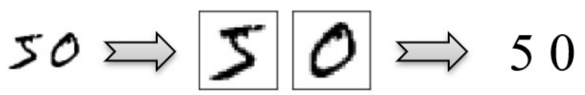

Fig. 1. Segmentation

4.2 Feature Extraction

Key features of the characters are extracted by applying a fixed threshold. Feature vectors are extracted which are used to distinguish a character from others.

In this paper, feature sets like the number of endpoints, number of junctions, horizontal projection count are calculated

4.3 Classification

Features gathered are used to identify the characters based on the decision-making classification algorithms. The feature vectors are matched with each selected class label.

The nearest neighbor classifier is one of the popular methods used for classification. The different types of classifiers that are popular are Support Vector Machine (SVM), Gaussian Naive Bayes, Decision Trees, Random Forest, K Nearest Neighbours, Stochastic Gradient Descent, etc. In terms of accuracy score, the SVM classifier was the most accurate, whereas Decision Trees are the least [27]. SVM technique is ideal to handle simple patterns. But as patterns get more complex then neural networks play an important role. Deep neural networks make use of more than 2 layers of the network to perform a task. Input features are fed to the input layer, weights are added and output is passed to the next layer. Weights are adjusted till the output is identified. Recurrent networks apply a feedback approach to provide dynamic artificial memory to machines. Long short-term memory (LSTM) is a special form of recurrent neural network and is efficient in terms of processing speed and predicting time series [27].

In this experiment, a neural network system is used for classification. Artificial intelligence (AI) empowers machines with the ability of human intelligence enabling machines to make decisions. The machine learning algorithms enable machines to deal with character recognition (CR) related problems. AI is an information processing system and consists of a large number of interconnected nodes which use trained artificial neurons in place of biological neurons. Weights are added to artificial processing nodes. After training the system with training algorithms, the system can perform like the human brain which can adjust to synaptic connections between the neurons. The machine is trained with a set of training data and weights are adjusted using training algorithms. Link weights store the knowledge required to solve the problem. Training algorithms provide the network with the ability to learn and adjust itself till it can recognize the input.

To recognize segmented digits, the network is divided into 3 layers - input, hidden, and output layer. If there are more hidden layers, it leads to a deep neural network or multilayer perceptron (MLP). In such cases, the output generated by a layer becomes the next layer's input. Such a network is referred to as a feed-forward network.

The functions of the three layers are:

Input layer - Input is fed in a grayscale format where 0 represents white and 1 represents black.

Hidden layer – weights are added and output is generated.
Output layer – predicts the output based on the highest activation value. (To predict digits from 0-9, 10 output neurons are used.) [3]. Fig 2., shows the neural approach to predict the digits 0-9.

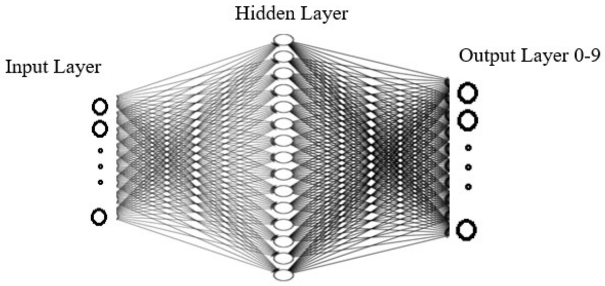

Fig. 2. Neural Approach to predict 0-9 digits

4.4 Recognition

The current era of ANN perceptron uses a function called ‘heaviside’ as an activation function. Heaviside function, H(z) = 0 for z<0 and 1 for z>=0. Some of the other frequently used activations functions for NN models are Hyperbolic tangent, Sigmoid, ReLu, Signum function, Softmax, etc

5. Implementation

Training data set is collected from MNIST and other samples which are not in the data set. These samples are scanned using a scanner or taken from a camera. The images are in the grayscale format of size 26X26 pixel. Some images from the MNIST data set are used to train or test data. Test data is used to check how well is the network system functioning at training the network.[3]. After loading the dataset, the training data is prepared so that it has the array size equal to the number of samples but with reduced dimensionality. The dataset is split into training and test data using the split function. The system is coded and algorithms are used to classify, make predictions, and find the accuracy rate.

Different outputs are generated for each classifier that gives the precision, accuracy, true-false positive values, and true-false negatives.

At each learning node, samples from our training set are presented. The error rate is calculated. The training process is continued till the error becomes less than a specified error limit. The network takes input, assigns weight and biases, and recognizes the character. It converts images to a form that is easier to process. The different stages of this model include input, convolution, pooling, fully interconnected, and softmax.

Pixel-level input is directly fed to the network. At a time only a portion of the input image is processed at each level.
The output of each operation is moving to the next level. Each operation helps in extracting the predominant features of the object. Features thus obtained are gathered together to form the feature set. The pooling layer reduces the image size [28].

To find the error(E)/loss/cost function, we use the following equation in our training algorithm:

$$E(wt,bi)\equiv 1/2n(\sum i\|y(i)-a\|^2) \qquad (1)$$

Here,

wt-weights

bi-bias

n-training input

i-input

a-output

By using equation (1) the error E(w,b) weights and bias are solved as minimum as possible. Also, to solve minimization issues gradient lowering technique is adapted. Gradient lowering or minimization algorithms are used to find weights and biases to minimize the error. To train the neural network, each time randomly small groups of training inputs are chosen and trained. This process continues till all training inputs are used.

$$wt_i \rightarrow wt'_i = wt_x - \eta/m(\sum j\partial CX_j\partial wt_i) \qquad (2)$$

$$bi_y \rightarrow bi'_y = b_y - \eta/m(\sum j\partial CX_j\partial bi_y) \qquad (3)$$

To keep the gradient to a minimum, the following update rule is applied repeatedly:

$$v \rightarrow v' = v - \eta\nabla E \qquad (4)$$

where E is a function of vector v with m variables. This rule equation (4) defines the gradient descent algorithm. The position of v is repetitively altered to obtain a minimum or least value for function (C). But this rule may not follow always apply as several other attributes can go erroneous and prevent gradient descent from obtaining the global minimum of C. Practically, it is found that gradient descent often works exceptionally well, and in neural networks, it paves powerful ways in terms of minimizing the cost function and facilitating the net learning capabilities.

6. Experimental Result

In this experiment, we mainly focused on handwritten digits. In this experiment, to train the network, 26x26 neurons are used in the first layer called input, n neurons are used in the subsequent layers (hidden), and ten output neurons. We have taken 10 output neurons to recognize the curves of the shapes accurately compared to lesser (say 4) output neurons. Hidden neurons can find the features of the digits.

The objective of this paper is to develop a system that can be used to sort out minimization issues and curtail error E(w,b). To perform the algebra, the Python shell program is used. To set up the network, MNIST data is loaded. This process is time-consuming but after training the network, the network will work faster. We have experimented with 30 neurons, learning rate $\eta=3.0$ in the hidden layer.

Recognition rate after each training is shown in Table 1.

Table 1: Recognition rate after each training

Epoch	***Recognition rate out of 10,000 test images***
Epoch 0	8112
Epoch 1	9055
Epoch 2	9134
Epoch 28	9545
Epoch 29	9512

Results show that the peak value of recognition rate is 95 percent for Epoch 28. The performance of the network improves over time.

To obtain better accuracies, better choices must be carried out in choosing the parameters - epochs of training, batch sizes, and learning rate. Better results are obtained by increasing the learning rates.

7. Conclusion

Handwritten recognition poses challenges to existing approaches for recognition tasks. To improve the accuracy rate, preprocessing techniques are carried out to eliminate the unwanted features in the image.

In this paper, we have presented a network system that can recognize digits accurately. Experimental results show that the more the network is trained with dataset sample, the more efficient is the system. The accuracy of the network chiefly depends on the selection of parameters like training epochs, batch sizes, and learning rate.

Reference

[1] Pai, A. "CNN vs. RNN vs. ANN—analyzing 3 types of neural networks in deep learning." Analytics Vidhya, Feb 17 (2020). https://www.analyticsvidhya.com.

[2] Wu, Jianxin. "Introduction to convolutional neural networks." National Key Lab for Novel Software Technology. Nanjing University. China 5, no. 23 (2017): 495. , https://cs.nju.edu.cn

[3] Learning, Deep. "Ian Goodfellow, Yoshua Bengio, Aaron Courville." The reference book for deep learning models (2016)., http://neuralnetworksanddeeplearning.com

[4] Sadouk, Lamyaa. "CNN approaches for time series classification." In Time Series Analysis-Data, Methods, and Applications, pp. 1-23. IntechOpen, 2019., DOI: 10.5772/intechopen.81170.

[5] LeCun, Yann, Bernhard Boser, John Denker, Donnie Henderson, Richard Howard, Wayne Hubbard, and Lawrence Jackel. "Handwritten digit recognition with a back-propagation network." Advances in neural information processing systems 2 (1989).

[6] Wan, Li, Matthew Zeiler, Sixin Zhang, Yann Le Cun, and Rob Fergus. "Regularization of neural networks using dropconnect." In International conference on machine learning, pp. 1058-1066. PMLR, 2013.

[7] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012): 1097-1105.

[8] Graves, Alex, and Navdeep Jaitly. "Towards end-to-end speech recognition with recurrent neural networks." In International conference on machine learning, pp. 1764-1772. PMLR, 2014.

[9] Sarikaya, Ruhi, Geoffrey E. Hinton, and Bhuvana Ramabhadran. "Deep belief nets for natural language call-routing." In 2011 IEEE International conference on acoustics, speech and signal processing (ICASSP), pp. 5680-5683. IEEE, 2011.

[10] Smolander, Johannes, Matthias Dehmer, and Frank Emmert-Streib. "Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders." FEBS open bio 9, no. 7 (2019): 1232-1248.

[11] Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." Psychological review 65, no. 6 (1958): 386.

[12] Riedmiller, Martin, and Heinrich Braun. "A direct adaptive method for faster backpropagation learning: The RPROP

algorithm." In IEEE international conference on neural networks, pp. 586-591. IEEE, 1993.

[13] Fitch, Frederic B. "Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. Bulletin of mathematical biophysics, vol. 5 (1943), pp. 115–133." The Journal of Symbolic Logic 9, no. 2 (1944): 49-50.

[14] Smalheiser, Neil R. "Walter pitts." Perspectives in Biology and Medicine 43, no. 2 (2000): 217-226.

[15] Hebb, Donald Olding. The organization of behavior: A neuropsychological theory. Psychology Press, 2005.

[16] Rosenblatt, Frank. The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory, 1957.

[17] Kuan, Chung-Ming, Kurt Hornik, and Halbert White. "A convergence result for learning in recurrent neural networks." Neural Computation 6, no. 3 (1994): 420-440.

[18] Forssell, Mats. "Hardware implementation of artificial neural networks." Information Flow in Networks 18 (2014): 1-4.

[19] Pradeep, J., E. Srinivasan, and S. Himavathi. "Diagonal based feature extraction for handwritten character recognition system using neural network." In 2011 3rd international conference on electronics computer technology, vol. 4, pp. 364-368. IEEE, 2011..

[20] Sharma, Om Prakash, M. K. Ghose, and Krishna Bikram Shah. "An improved zone based hybrid feature extraction model for handwritten alphabets recognition using euler number." International Journal of Soft Computing and Engineering 2, no. 2 (2012): 504-508.

[21] Suen, C. Y. "Character Recognition by Computer and Applications in Handbook of Pattern Recognition and Image Processing, ed. Young TY, Fu, KS." (1986): 569-586.

[22] Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3128-3137. 2015.

[23] Bunke, Horst, Samy Bengio, and Alessandro Vinciarelli. "Offline recognition of unconstrained handwritten texts using HMMs and statistical language models." IEEE transactions on Pattern analysis and Machine intelligence 26, no. 6 (2004): 709-720.

[24] Kozielski, Michał, Patrick Doetsch, and Hermann Ney. "Improvements in rwth's system for off-line handwriting recognition." In 2013 12th International Conference on Document Analysis and Recognition, pp. 935-939. IEEE, 2013.

[25] Bluche, Théodore, Hermann Ney, and Christopher Kermorvant. "Feature extraction with convolutional neural networks for handwritten word recognition." In 2013 12th International Conference on Document Analysis and Recognition, pp. 285-289. IEEE, 2013.

[26] Vivek Singh Bawa, https://pydeeplearning.weebly.com/blog/basic-architecture-of-rnn-and-lstm, Basic architecture of rnn and lstm 1/18/2017, blog.

[27] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9, no. 8 (1997): 1735-1780.

[28] Prabhu, Raghav. "Understanding of convolutional neural network (CNN)—deep learning." Medium. Com (2018): 1-11.

Ashraf Ali received his B.Sc. (Computer Science and Mathematics) from University of Lucknow, India, Master of Computer Applications from Bundelkhand University, India, and Ph.D. in Computer Science and Engineering from Singhania University, India. He is currently working as an Assistant Professor at the Faculty of Computer Studies, Arab Open University, Kingdom of Bahrain. His research interest includes Cloud Computing, Web Mining, Web Information Retrieval, Semantic Web, Semantic Similarity, E-Learning, and Artificial Intelligence. He has published numbers of articles in the various journal of repute.

Anita Venugopal received her Ph.D. in Computer Science and Mathematics from Motherhood University, India. She is currently working as Lecturer at Dhofar University, Oman. Her research interest includes Artificial Intelligence, Database systems, Machine learning and Distributed Computing. She has published number of research papers in various referred International/ National Journals and Conferences.