Artificial Neural Networks

Introduction

Artificial intelligence is one of the most popular computer science technology in these 10 years. Google, Microsoft and IBMI also invest amount of money in the research of artificial intelligence. Among this, it is believed that machine learning is an important research field. To make the machine intelligent, artificial neural network is regarded as an efficient way. Artificial neural network is used in many different field to solve problems, such as electric load forecasting [1], clinical medicine [2], financial and economic forecasting [3] and so on. Artificial neural networks are also closely related to deep learning, for example image identification. Artificial neural network makes considerable outcome, which appeals people pay more attention to this field.

In this report, firstly, the brief introduction of machine learning, artificial neural network and backpropagation algorithms is given. Then the practical of creating an appropriate artificial neural network for “Guess who” game is implemented. To illustrate the process of building this neural network efficiently. This report includes the design steps of part1 and part 2 requirements, examples of the main methodologies and the approach of testing. Furthermore, the evaluation of the artificial neural network is analysed. Last but not least, the running instructions of this system is displayed.

1. Literature review

1.1 Machine learning

One of the most important approaches of artificial intelligence is machine learning. Machine learning progress promoted the advances in many fields, such as computer vision, language processing, robotics and many other fields, which enhances its status in artificial intelligence [1] Machine learning can reflect intelligent behaviour. It is regarded as an area of improving computational theories of learning processes and creating learning machine [2].In Machine Learning: A probabilistic Perspective, machine learning is defined as “a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty” [3] At present, the objectives of Machine learning mainly centralize on task-oriented studies, cognitive simulation and theoretical analysis [4]. Machine learning methods includes supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. Machine learning can make machine adapt to new environment and speculate results [5]. Creating artificial neural network is a common way to realize machine learning.

1.2 Artificial neural network

The existence of human brain demonstrates that biological neural networks realize human cognition, perception and behaviour control [6]. Artificial neural network originates from imitating the mechanism of nervous system of human brains [7]. The fundamental elements of artificial neural networks are neurons, or nodes or

units, which is similar to cells in human nervous system [8]. In artificial neural network model, there is an input layer, an output layer and various hidden year consist by artificial neurons with the signal being transmitted from input layer to hidden layer to output layer [9]. Although the direction of information flow is fixed, but artificial neural networks are able to be improved through justifying the weight of connections, the activation function and the number of nodes or hidden layers based on feedback mechanism of neural network [10]. When training or testing neural networks, the choice of the number of neurons, the initial weights and the learning rate are effectively influence the quality of neural network [11].

1.3 Back-propagation training

Feedback neural networks are widely used in different field. Among it, the backpropagation is a common method to fit feedback neural networks [12]. Back propagation algorithms are typically used in supervised learning [13]. Common backpropagation algorithms include Levenberg–Marquardt, Rprop, DeltaBar-Bar, Quickprop, Vogl and SuperSAB algorithms [14]. Backpropagation has the disadvantages of: requiring many iterations to converge and therefore inappropriate for on- line real-time learning (this imposes the need to use small networks, and therefore many difficult problems of current interest cannot be solved [15]. One of the well-known strengths of back propagation is that it will gradually assign zero weights to inputs which provide no information [16].

2. Design

This practical is to construct an artificial neural network to implement the simplified “Guess Who” game. “Guess Who” is a two-player character guessing game. In the simplified game, there are 5 characters, and every character has their 7 typical features. According to these features being asked. The player guesses the character of the image that the opponent holds.

Some material has been given including a table file (char.csv) with 35 groups’ data of “Guess Who” game. Every group consists of answers of 7 features and the corresponding character name. Based on this table, appropriate neural network to guess the opponent character is asked to be built.

2.1 Part 1

The aim of part 1 is to build an appropriate neural network for single player. The agent asks player to answer questions, and after asking 7 features, the agent give a guessing result. The guessing result is expected to be as correct as possible, which basically means the neural network is appropriate. The following content is what I did for every point.

2.1.1 Point1

The file “char.csv” provides 7 features for each person. According to these 7 features, the artificial intelligence agent judges the character. Therefore, there are 7 inputs in this practical.

Common encoding methods contains one-hot, one-of-N and binary. One-hot and one-of-N is usually used to code a small number of inputs or outputs but binary can be used in coding a large quantity of input or output. Additionally, in this situation, it is required to make sigmoid function as activation function, which means one-of-N does not satisfy the requirement. The reason is that one-of-N return zero after normalisation. Last but not least, one-hot encoding is safer than binary when are used in simple neural network. Thus, nodes in input and output layers use one-hot encoding methods.

As the rules of one-hot encoding method, after calculating, the number of output should be 5.

2.1.2 Point2

In this step, firstly the file containing dataset is read. After that the method of “createTrainingTable” is used to transfer the content of dataset into a training table in an appropriate format. In this method the one-dimensional array named paragraphs is divided to two-dimensional array called table. There are two “for” loops in the method. Namely, one is for splitting up the elements of paragraphs array, and another is for converting “Yes” to “1” and “No” to ”0”.

2.1.3 Point3

This network includes one hidden layer. The number of the nodes in hidden layer can between 3 to 10. After justifying the number of nodes and checking the corresponding graphic, 5 nodes of hidden layer are chosen.

2.1.4 Point4

There is a single hidden layer in this neural network, the number of neuros in this layer is designed to 5 where the line chart is more smooth and appropriate. There are three usual formula to calculate the initial number of nodes in hidden layer. First, the quantity of nodes in hidden layer can be between the number of nodes of input layer and the number of nodes of output layer. Second it can be lesser or equal to the number of units of input layer. Last method seems more complicated which make it more than the sum of the two thirds times of the number of units in input layer and the number of the units in output layer. Based on these three approaches, the number of hidden layer can be tried from 5 to 10.

2.1.5 Point5

After training the network using backpropagation, the learning rate and training set are set still as 0.5.

2.1.6 Point6

After changing the number of hidden units, adjusting appropriate learning rate and momentum, the current perfect neural network need to be saved. In this stage, “saveObject" in "EncogDirectoryPersistence" is used to save the neural network.

2.1.7 Point7

Use the BasicMLData function and compute function to obtain inputs and calculate outputs. Due to the threshold is 0.5, the outputs can be coded by 1 and 0. Then compare the output with real output.

2.1.8 Point8

Because of the one-hot coding method is used, the pruning method not needs to be used. After checking the diagram with different number of hidden units. The number of 5 is chosen. It's no fluctuation and decreases properly, therefore it is appropriate.

2.2 Part 2

This part is the enhancement of part 2. The neural network built by part 1 is loaded in part 2 to test if it is efficient in "Guess who" game. In addition, the system should be able to guess the result in advance. When there are fewer number of answers being asked, the system still can give an early but relatively reliable guess.

2.2.1 Point1

As it shown in the file (char.csv), there are 7 features. To correspond to the dataset, the sequence of questioning should be in same order. In this code, features are stored in an array. One feature is extracted from array each time through" for" loop when the player is prompted to answer "Yes" or "No" to the agent. Storing feature into an array is more convenient, brief and flexible if other features need to be added.

2.2.2 Point2

In this step, the neural network from part 1 is loaded through load method followed by computing the output of the given input pattern

2.2.3 Point3

When the user enters a pattern without being included in the training set, the output calculated by the neural network may not be the format of one-hot. To solve this problem, I make the biggest number as 1 and make the rest numbers as 0. According to this algorithm, the guessing result is given.

2.2.4 Point4

At the end of getting all answers from the player, the system obtains the guessing result through convert the output given by the agent to the name of corresponding character. The output consists of 5 double type real number. After checking each number each time through "for" loop, if it is higher than 0.5, it is regarded as 1, otherwise it is regarded as 0. Then the system compares processed output with name codes to transfer processed output to name, in other words guessing result.

2.2.5 Point5

To add a new character and a new feature to exist training table and train a new neural network, the first thing I did is to build a new training table which includes previous information, a new feature and a new character. Then according to the new training table, a new network is created with changing the number of input layer and output layer. After justifying parameters and invoking “saveObject” function, a relative appropriate new neural network is saved. When the eighth question is answered, this new neural network is used to calculate the output.

3. Examples and Testing

3.1 Part1:

In part 1, there are two classes are built. Among them, there are 5 main functionalities being. Firstly, the class named FileUtil is used to read the content of char.csv file, which utilizes FileReader and BufferedReader class of java API. After obtaining the string array of the content of the file, it need to be transferred to two-dimensional training table in the method of createTrainingTable as preparation of creating artificial neural network. Then the neural network can be created by using of BasicNetwork class. In this process, the nodes of input layer, output layer and hidden layer should be ensured. Last but not least, after finishing all preparation, the nerual network can be trained by using of backpropagation algorithm until the error is smaller than 0.01. It is worth to noticing that the training result of the neural network should be saved by saveObject function. At the end of this process, the functionality of computing output is used to check if the guessing output is same as the output of training set.

3.2 Part 2:

Compared with part 1, it seems more complicated in part 2 because it is required to stimulate the process of simplified “Guess who” game. There are 3 classes in part 2 including 10main functionalities being applied. Among them, there are 5 functionalities are fairly similar with part 1. The major things have been changed is that a new training set with 8 features and 6 characters has been read, which also means the numbers of units of input layer, output layer and hidden layer has changed. However, these 5 functionalities are fixed.

Another 5 functions are used to realise the game. To begin with, the system prompts the user to input answers of questions and answers should be gained from keyboard by user typing. Secondly, 5 different trained and saved neural networks can be called according to the instructions of the number of acquired questions. For example, after the fourth question is responded, the neural network with 4 inputs will be called. Thirdly, after the seventh question is responded, the user will be asked that If want to add a feature and a character. If the response is no, the programme stops. If the response is yes, the programme continues to ask the eight questions. Then, whenever a neural network is called, the corresponding functionality that calculating input result and output result is executed. Finally, whenever a neural network is call, the value of output is converted to guessing result, namely, the character name is displayed.

4. Evaluation

4.1 Coding input and output

There are 3 common coding methods of input and output including one-hot, one-of-N and binary. One-hot and one-of-N are suitable for simple neural networks, which means the complicity of the neural networks are low. It can have represented by a few number of nodes of layers, a few number of hidden layers and so on. However, binary is suitable for complicated neural networks because it can has fewer number of output. One-hot and binary can be used in sigmoid activation function neural networks but one-of-N is not satisfied this kind of networks. The reason is that the graphic of sigmoid activation function is above x axis. In this practical, one-hot coding method is used because is safer than binary and it meet the demand of sigmoid activation function applied in hidden layer.

4.2 Bias

Normally, artificial neural networks have bias to make it general and more approach to real word. Bias nodes increase the flexibility of the neural network and it usually used in feed-forward networks. In this practical, the networks are relatively simple, therefore only one bias is added in hidden layer to improve the accuracy and flexibility.

4.3 The number of nodes of hidden layer

There are 5 networks in this practical. When the input is smaller than 7, the number of nodes of hidden layer is 3. When the input is equal to 7, the number of nodes of hidden layer is 5. When the input is equal to 8, the number of nodes of hidden layer is 6. Because of the simplicity of the network, the quantity of the units in hidden layer have no big influence on rationality of neural network. Take 7 input nodes for example, both graphic 1 and 2 are satisfied. However, it is commonly believed that smooth line is better.

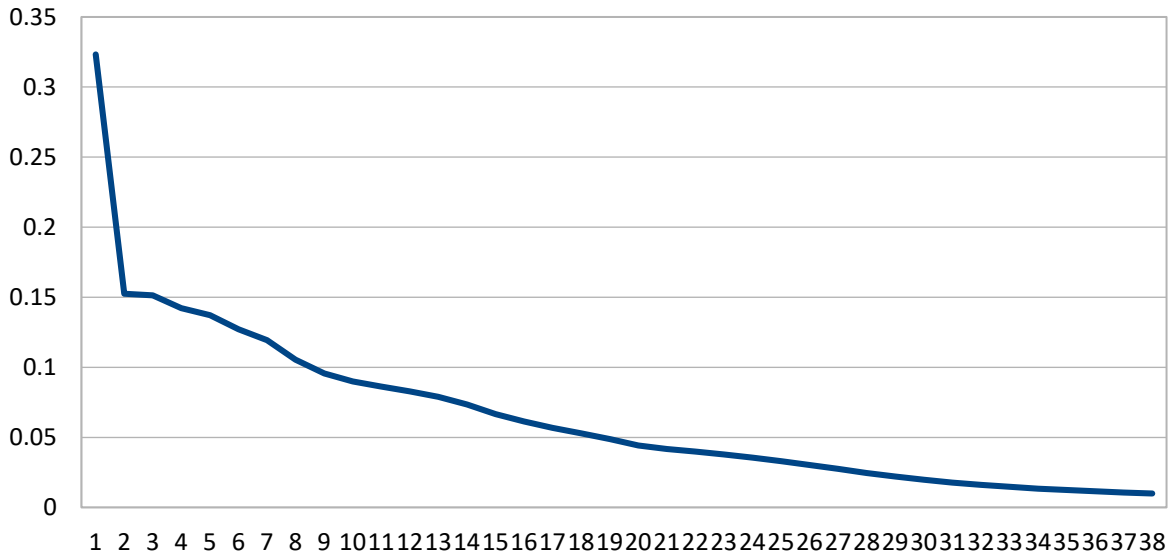

Figure 1 3 hidden units

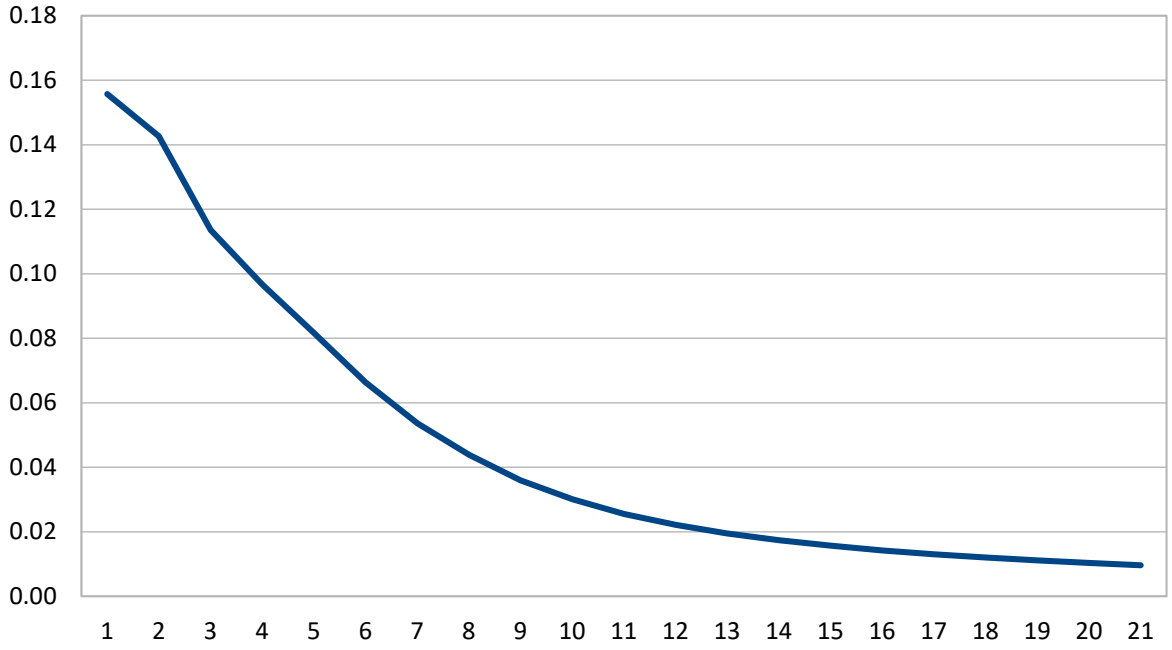

Figure 2 5 hidden units

4.4 The method of early guessing

To achieve a proper and accurate early guessing, the most important thing could be creating a proper network. Although I thought it is possible to assign 0.5 to the questions have not been asked, I suppose the accuracy and rationality of this method is unwarrantable. This because how 0.5 influences output calculating is unsure. Hence, to be more targeted and specific. Early guessing is implemented by different networks. For instance, if 4 questions have been asked, the 4 inputs neural network are used. Before the 4 questions, this system does not give early guessing because the randomness is so high. However, there is also another problem that when the number of questions have been asked is 4, the fact is that the error of 4 inputs neural network cannot be lower than 0.01. Thus, I changed the error to 0.02. Even though it cannot be regarded as a successful neural network, it still has a certain extent accuracy. Therefore these 4 inputs neural network is still reserved.

4.5 The build of new training table

As the requirement of part 2, a new feature and a new character should be added. Before doing that, it is better to analyse the content of original data set. There are 7 features and 5 characters. Each character has 7 data set, and the error rate is about 1/7~2/7 each time. Therefore, in this new training table, the sixth character and the eighth features are added. Meanwhile, each character has 8 data set now. It means because of the emergence of the eighth feature, every old character adds a new data set with the eighth feature. At the end, there are 48 rows and 8 column inputs data.

5. Running

5.1 Part 1：

The name of the file is learing1. The code is implemented followed by the steps of the requirement of part 1. In other words, when the program run, the error of a neural network is shown and the test result (as the step 7 in part 1) is demonstrated.

5.2 Part 2:

The name of the file is learnng2. A simplified “Guess who” game is programmed. There are 6 instructions given to run this system.

1. Open the char.csv and char_.csv
2. Run the program
3. The system asks you questions about features of the character. Type “Yes” if the character has the feature and type “No” if the character does not have the feature. Both “Yes” and “No” should be capital in initial letter and without space after the word.
4. If you only decide to answer the first seven questions, the file (char.csv) should be referenced to answer questions, and you should type “No” after the system ask you “would you like to add a character and feature?”.
5. If you decide to try the eighth question, you should open char_.csv file and answer questions based on the dataset of this file, and you type “Yes” after the system ask you “would you like to add a character and feature?”.
6. Notice that when you want a new character and a new feature, namely you decide to try the eighth question, you should ignore the early guessing because that early guessing is only for 7 inputs and 5 outputs neural network instead of for 8 inputs and 6 outputs neural network.

6. Reference

[1] Russell S, Norvig P, Intelligence A. A modern approach. Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs. 1995; 25:27.

[2]. Kodratoff Y, Michalski RS. Machine learning: an artificial intelligence approach. Morgan Kaufmann; 2014 Jun 28.

[3] Murphy KP. Machine learning: a probabilistic perspective. MIT press; 2012 Sep 7

[4] Michalski RS, Carbonell JG, Mitchell TM, editors. Machine learning: An artificial intelligence approach. Springer Science & Business Media; 2013 Apr 17.

[5] Nilsson NJ. Artificial intelligence: A modern approach: Stuart Russell and Peter Norvig,(Prentice Hall, Englewood Cliffs, NJ, 1995); xxviii+ 932 pages.

[6] Yegnanarayana B. Artificial neural networks. PHI Learning Pvt. Ltd.; 2009 Jan 14.

[7] Zahedi F. An introduction to neural networks and a comparison with artificial intelligence and expert systems. Interfaces. 1991 Apr;21(2):25-38.

[8] Mehrotra K, Mohan CK, Ranka S. Elements of artificial neural networks. MIT press; 1997.

[9] Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods. 2000 Dec 1;43(1):3-1.

[10] Freeman JA, Skapura DM. Algorithms, Applications, and Programming Techniques. Addison-Wesley Publishing Company, USA; 1991.

[11] Dayhoff JE, DeLeo JM. Artificial neural networks. Cancer. 2001 Apr 15;91(S8):1615-35.

[12] Yu, X., Onder Efe, M. and Kaynak, O.: A general backpropagation algorithm for feedforward neural networks learning, IEEE Transactions on Neural Networks 13 (2002), 251–259.

[13] Riedmiller M. Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms. Computer Standards & Interfaces. 1994 Jul 1;16(3):265-78.

[14] Hagan, M. T. and Menhaj, M. B.: Training feedforward neural networks with the Marquardt algorithm, IEEE Transactions on Neural Networks 5 (1994), 989–993.

[15] Miller WT, Glanz FH, Kraft LG. Cmas: An associative neural network alternative to backpropagation. Proceedings of the IEEE. 1990 Oct;78(10):1561-7.

[16] Specht DF, Shapiro PD. Generalization accuracy of probabilistic neural networks compared with backpropagation networks. InNeural Networks, 1991., IJCNN-91-Seattle International Joint Conference on 1991 Jul 8 (Vol. 1, pp. 887-892). IEEE.