平成 28 年度　博士学位論文

主論文

Forecasting Time-series Using Statistical Models and Neural Network

(統計モデルとニューラルネットワークを用いた時序列の予測研究)

理工学教育部

数理・ヒューマンシステム科学専攻

平成 26 年度入学

氏　名　　　　　虞　瑩

学籍番号　　31471006

指導教官　　　唐　政

Table of contents

List of Tables

List of Figures

Abstract

Time-series forecasting techniques have been applied in some real-world applications, for example economic environment and market forecasting of finance, electric utility load forecasting, weather prediction and reliability forecasting. Methods of forecasting time-series have attracted more and more attention of both academics and practitioners. However, there has not been a panacea for each time-series.

Time-series is an accumulation of data points and is accumulated in a fixed time interval. And it is intend for analyzing and predicting long-term trend. Then the future can be forecasted and it can also perform some other form of analysis. There are two things make a time-series different from a regular regression problem. One thing is time dependent. The other thing is accompany with an increasing or decreasing trend, and almost all of the time-series have one or more form of seasonality trends. The data set can be loaded by many ways, and so as to load the data as a time-series data, some special arguments should be passed, such as differential, moving average and other ways. The time series is stationary if its statistical properties remain unchanged: the mean and the variance remain unchanged with time. However, why can it be said important? This is because most of the time-series models are on the basis of assumptions that the time-series is stationary. Obviously, we can express the opinion that if a time-series has a particular behavior over time, it can be highly represented that it will behave the same trend in the future. Similarly, the quiescent series of theories focuses on more mature and easier to implement than stationary sequences. Therefore, the use of static is very strict standard definition. And time-series can be considered to be stationary if the statistical properties are sustained over a long period of time: 1. Constant mean; 2. Constant variance; 3. Auto-covariance that does not depend on time. Although the stationary hypothesis is used in many time-series models, and almost no actual time-series is stationary. For that, Statisticians have found a series of methods fixed. In fact, it is almost impossible to completely stop the time-series, but I can do my best to make it as much as possible by adjusting the trend and seasonality in my paper.

Up to now, there are many researchers using a lot of methods to forecast the

time-series. And they can be divided into three types: statistical models, neural network and combined models. The most popular statistical models are the Naïve model, the exponential smoothing model (ES), and the autoregressive integrated moving averages model (ARIMA). They are almost used to forecast the linear models. Recently, more and more nonlinear forecasting models are proposed to address the time-series' issues. Among them, artificial neural networks (ANNs) are receiving increasing interests due to their ability to imperfect data, functions of self-organizing, self-study, data-driven, associated memory, and arbiter function mapping. As we all know, the structure of every neuron is unique, it contains three components: the cell body, dendrite and axon. The dendrite receives signals from other neurons, then the signal is computed at the synapse and transmitted to the cell body. If the signal which was sent into the cell body exceeds the holding threshold, the cell will fire and send signals to other neurons by the axons. Follow this theory, the DNM model (Dendritic Neuron Model) was proposed in this thesis. Both linear and non-linear models have achieved great success in their own linear or nonlinear case. But none of them is an omnipotent model for all situations. So I can assume that a combined model of linear and nonlinear modeling capabilities may be a perfect choice for the prediction of time-series data. The purpose of my research is to find a super model which can forecast any time-series precisely.

Acknowledgements

I very sincerely thank my Professor Zheng Tang, and my Associate Professor Shangce Gao for their suggestions and support during the research. In the whole course of my research, I benefited from their advice and encouragement. It was a great pleasure to me for conducting this thesis under their supervision.

I also want to thank members of Tang Lab. Thanks for sharing the knowledge in the research field with me, giving me some good advices in my research and discussing with me about some theories on the field of neural network and algorithm optimization.

Of course, I am grateful to my parents. Without their love and patience, my research would never have come into existence. I am also very grateful for my husband Shuya Ding whose love, patience and encouragement during the PhD period enabled me to complete my research.

Finally, I want to thank all the people who gave me kindly assistant and consideration during the period of my PhD.

1. Introduction

1.1 Introduction of Time-Series

Time-series is an accumulation of data points and is accumulated at constant time intervals. Time-series forecasting techniques have been applied in some real-world applications, for example economic environment and market forecasting of finance, electric utility load forecasting, weather prediction and reliability forecasting. The approaches of time-series analysis precedes the general stochastic process and Markov chain.[1],[2] The purpose of time-series analysis is describing and summarizing time-series data, fitting the models of low-dimensional, and making the predictions.

When analyze a time-series data, the following steps should be gone through: [3],[4]

① What lets time-series data particular?
② How to loading and processing the time-series data.
③ How to examine smoothness of a time-series data?
④ How to turn a time-series data to be smooth?
⑤ Prediction of time-series data.

① What lets time-series data particular?

As its name suggests, the time-series is a collection of data points at the fixed time intervals. It is used to analyze and determine the long-term trend, then it can be used to predict the future or perform some other form of analysis. However what lets a time series data special, so it can be said that it is unlike ordinary regression problems? There are two things:

- One thing is the time dependent. It is determined by the basic hypothesis of a linear regression model that underlying observations are independent and not held.
- The other thing is that it is along with an increasing or decreasing trend, There are some forms of seasonal trends in almost every

time-series data. For example, changes which are special to a specific time frame.

Due to the inherent nature of time-series data, there are different kinds of steps to analyze it. They will be discussed in detail as below.

② How to loading and processing the time-series data.

The data set can be loaded by many ways, and for reading the data as a time-series, some special arguments should be passed, such as differential, moving average and other ways. There are 2 things to be noted. You should know the type of the data before you analyze it, whether it is numeric or string or some other else. Then the range of data will be sorted for working. If you confuse the data randomly, it will not work.

③ How to examine smoothness of a time-series data?

The time-series can be suggested to be smoothness when its statistical properties such as mean, variance can keep fixed over time. However why is this important? Almost every time-series model is based on the hypothesis that the time-series data is smoothness. Intuitively, it can be said that when a time- series has a similar behavior over a long period of time, there will be a very high probability that it will follow the same trend in the future. Similarly, the fixed time-series correlation theory is more mature and simpler to implement than non-stationary series.

Smoothness definition has a very strict standard. However, we can assume the time-series to be smoothness for practical purposes if the statistics persist over a long period of time, as the follow shown:

- Whether it has a constant mean or not.
- Whether it has a constant variance or not.
- Whether there is an auto-covariance that does not depend on time.

How to test smoothness of time-series. The first thing is to plot the data simply and analyze visually. It will clearly evident trends in the data.

However, it might not always be possible to make such visual inferences. So, more formally, we can check the time-series is smoothness or not by using the following methods:

- **Rolling Statistics.** It means that we can draw the moving average or moving variance and check whether it changes as the time changes.
- **Dickey-Fuller Test.** Dickey-Fuller is a statistical test which can be used to check the data is smoothness or not.

④ How to turn a time-series data to be smooth?

As we all known, almost every time-series models is in the view of the assumption of smoothness. However, there is no practical time-series is smoothness. For this reason, statisticians have already found ways to stabilize time-series. In fact, it is obviously impossible to make a time-series completely static, but we can try to make it as much as possible.

First we should have an opinion that what lets a time-series to be not smooth. There are two major reasons as follow:

- The first one is the trend, for it always varies mean over time.

- The second one is the seasonality, because of the variations at specific time-frames.

The fundamental principle is to model or estimate the trend and seasonal ability of the time-series and to remove the trend and seasonal time-series of stationary time-series. Statistical prediction techniques or other approaches can then be used to implement the time-series. The final step is to convert the predicted value to the original size by applying the trend and seasonal constraints.

⑤ Prediction of time-series data

We can use a simple method to describe a time-series which we call it the classical decomposition. The viewpoint of it can be concluded that the time-series could be divided by 4 elements:

- First one is the trend (Tt) ,it means thelong term movements in the mean;
- Second one is the seasonality (It), because the cyclical fluctuations are thought to have connection with the calendar;
- Third one is the cycles (Ct), for example, some else cyclical fluctuations (such as a business cycles, and so on);
- Last is the residuals (Et), it means other random or perhaps we can call it the systematic fluctuations.

The purpose is to bring about the separate model which is determined by the 4 elements, after that, we can combine them together, which can be shown as $Xt = Tt + It + Ct + Et$ or $Xt = Tt \cdot It \cdot Ct \cdot Et$ by using multiplication.

We have seen different approaches and each of them worked reasonably well for letting the time-series stoothness. Up to now, there are many researchers using a lot of methods to forecast the time-series. And they can be divided into three types: statistical models, neural network models and combined models. The most popular statistical models are the Naïve model, the exponential smoothing model (ES), and the autoregressive integrated moving averages model (ARIMA). They are almost used to forecast the linear models. Recently, more and more nonlinear forecasting models are proposed to address the time-series' issues. Among them, artificial neural networks (ANNs) are receiving increasing interests due to their ability to imperfect data, functions of self-organizing, self-study, data-driven, associated memory, and arbiter function mapping. In my paper, I will introduce several statistical models and several neural networks models to fit and forecast the time-series data, and compare the effective of them. I will also use the inbound tourism data of Japan and the data of Chinese house price index to test the models I proposed.

1.2 Introduction of Statistics

1.2.1 Basic of statistics

Statistics is a science of a very wide range of subject, which is applied in a large number of different territories. Under normal circumstances it can be said that statistics is the technology of collecting the data in order to analyze it, after interpreting it, conclusions can be drawn from information through the previous steps. Another point of view, statistics is the technology that scientists and mathematicians have created to interpret and summarize conclusions from the information by collecting and analyzing the data. Everything that deals with the process of collecting, analyzing, interpreting and presenting of data is belong to the process of statistics, including all detailed planning and prior activities.[5]

As mentioned above, it is clear that statistics is not just data collection and drawing the characterization of the data collected, but on the basis of these, statistics can classify the data and analyze it, it is a science which can provide answers to the following questions:

- How much of the determined kind of data should be collected?
- How could we range and summarize the data that we collected?
- How could we analyze the data and give the conclusions after analyzing it?
- How could we estimate the effectiveness of the conclusions and how could we assess their uncertainty?

For that, statistics provides approaches for those:

① For the design, which can plan and carry out research studies.
② For the description, which can summarizing and exploring data.
③ For the inference, which can make forecasts then summarize the possible phenomena that is represented by the data.

In addition, statistical data are available to deal with uncertain scientific phenomena and events. It is successfully applied in different areas of research. Certainly, statistics are now used in all areas of science. [6]

Statistics is a science with 2 basic concepts: Population and sample. Population which is defined as a group of individuals or object investigators which are primarily interested in their research questions. In some cases, measurements of all individuals desired in a population are obtained, but in most cases, only a set of subjects of an individual or group is observed, such a set of individuals that may constitute a sample. There are always only certain, relatively few, characteristics of individuals or objects that are investigated at the same time. Not all attributes need to be measured from a population of individuals. This observation emphasizes the importance of a set of measurements, so it gives us an alternative definition of population and sample.

The population tends to represent the objective of a survey. We can understand the collection of samples that are generally sampled. There may be many different populations, the difference of population and sample is shown as Fig.1.1.

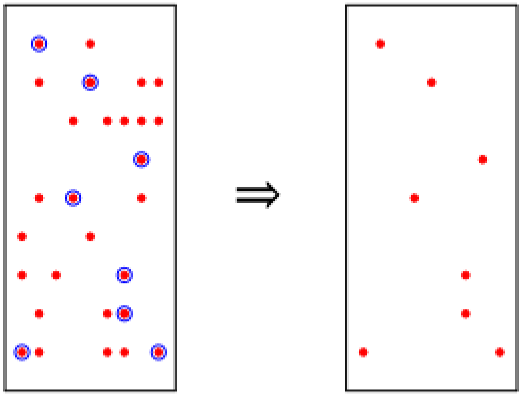


Fig.1.1 Population & Sample

Statistics has two major types called descriptive statistics and inferential statistics. The embranchment of statistics which is mainly used in the description and summarization of the data is suggested as descriptive statistics. While the embranchment of statistics related with utilizing sample data to make inferences about the population that is suggested as inferential statistics.

Descriptive statistics contains not only the construction of graphs, charts, and tables, but also the algorithm of different kinds of descriptive measures. For instance, the mean, measures of variation, and percentiles.

Inferential statistics includes the methods such as point estimation, interval estimation and hypothesis testing. They are all built on top of the probability theory.

Descriptive statistics is interrelated to inferential statistics. It is always necessary to utilize approaches of descriptive statistics to describe and summarize the information which is included in the sample, after doing these, approaches of inferential statistics could be utilized to make more comprehensive analysis of the problem under investigation. In addition, the preliminary descriptive analysis of the sample shows the characteristics of these characteristics which are often the result of the appropriate inference methods chosen for use.

As we all known, the data can sometimes be collected throughout the population. In this case you can perform a descriptive study on the population as well as on a sample. Only when reasonably draws the conclusion based on the population derived from the information obtained from the sample inference.

1.2.2 Describing Distributions with Numbers

As Andrejs Dunkels said: It is easy to lie with statistics. However it is also hard to tell the truth without it. Statistics can be very important to the truth. When analyzing the observations, we will observe and analyze the two basic attributes of the observations. One is a measure of the center value which is calculated by the data, for example, the median and the mean. In connection with this measurement, we can add a second value that describes how these observations propagate this given central measurement. The median is the central observation of the data. After sorting from the lowest observed value to the highest observed value [7]. In addition, in order to give a sense of expansion in the data, we often give the minimum and maximum observations and observations that are 1/4 and 3/4 of the rise of the list, which is known in the list Of the first and third quartiles. For this average, we usually use the standard deviation to represent the spread of the data. In measuring spread, five number summary is always been used. The first quartile Q1 is the median of the lower half, and third quartile Q3 means the upper half of the data. The five-digit summary of the data represents the values of the minimum value (Q1), the median value (Q3), and the maximum value. These values, as well as the mean can be summarized in box-plot as Fig.1.2 shown.

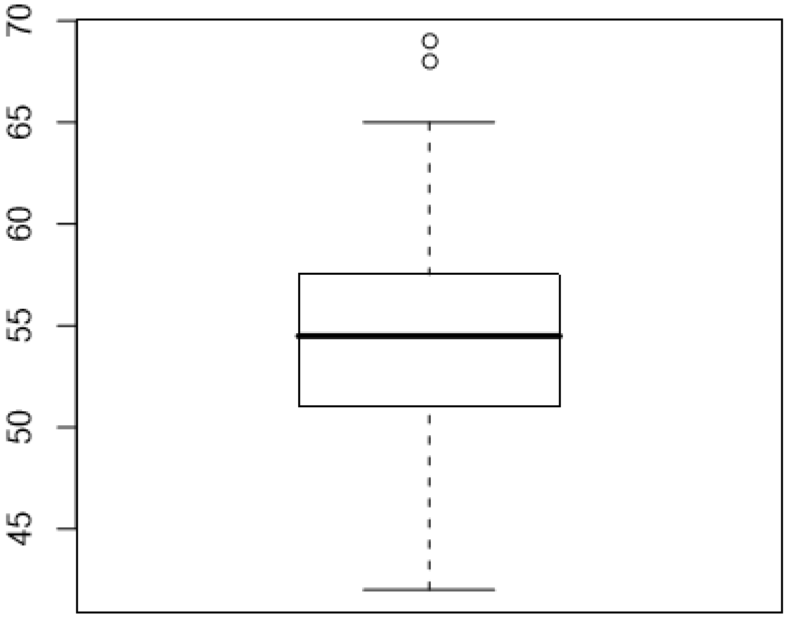


Fig.1.2 Example of the box-plot

1.2.3 Statistical data analysis

The purpose of the statistics is to make the data understandable. Any data analysis should include the following steps as Fig.1.3 shown:

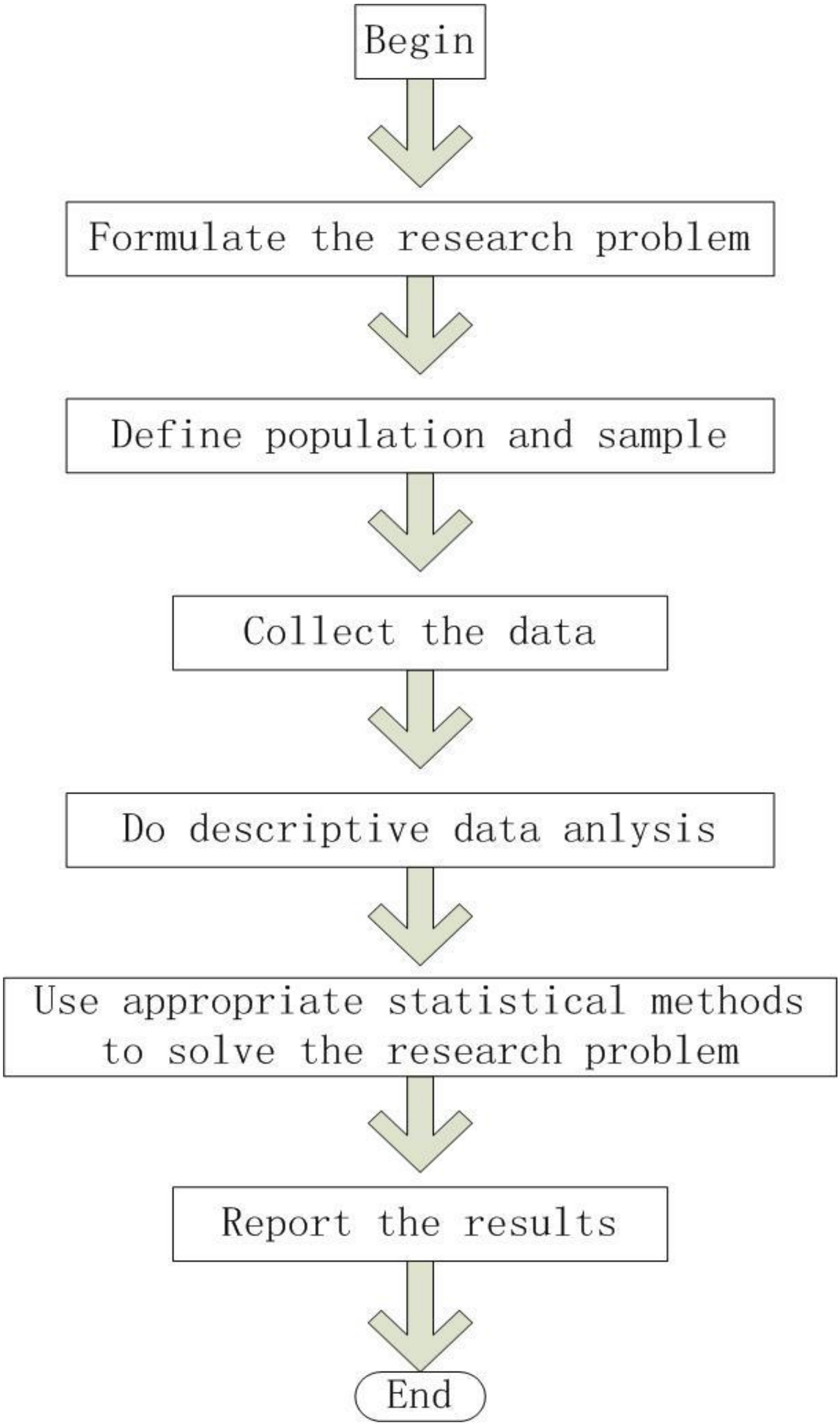


Fig.1.3 Steps of data analysis

To summary this section, we can discover that the major purpose of statistics is to extrapolate the population by analyzing the information contained in the sample data. This includes assessing the degree of uncertainty involved in these inferences.

1.3 Introduction of Neural Networks

The word "Neural networks" is very popularly mentioned in recent years. It means machines that something of them is like the brain, may be filled with science fiction mythology as the sci-fi content. [8] Neural networks are prone to sub-elements, units or nodes of the interrelated part of its function based on animal neurons and quickly. The branching capability of the networks are stored in the cross-connect strength, otherwise, through the weights which is contained by training by adapting the training pattern set or learning from the training pattern set.

In this section, we first introduce a relationship between biological neural network and artificial neural network, then describe a historical overview of artificial neural network. Finally, we will discuss the common problems in these neural networks when using them to solve combinatorial optimization problems.

1.3.1 Biological neural network to ANN

The brain is constituted of about 10 billion neurons for the most parts, each one is concerned to about 10,000 other neurons. Every neuron can receive electrochemical inputs, and the inputs are from other neurons at the dendrites as shown in Fig.1.4. When the summation of the electrical inputs has the plenty energetic to operative the neuron, it will transmit electrochemical signals through the axon, then pass the signals to another neuron whose dendrites attach to other neurons at the end of any axon.[9][10]

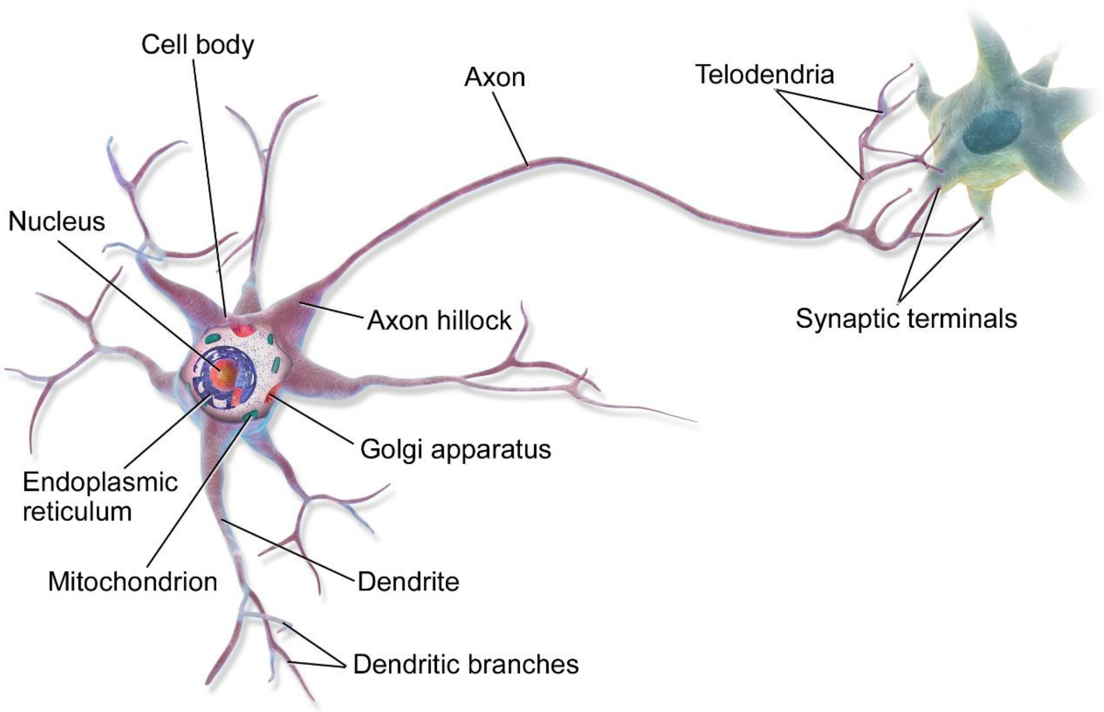


Fig.1.4 Structure of a typical biological
(neuronhttps://en.wikipedia.org/wiki/Neuron)

We should pay attention to note that the neurons will only be excited if the summarized signal exceeds a given level. It is used to decide whether neurons will emit or not.

Every neuron contains 4 main areas connected with its structure. The cell body, or we can call it soma which contains the nucleus and maintains the protein mixture. Neurons also have some dendrites, and the branch like a tree-structure. However, neurons usually have just one axon, it grows from a part of the cell body, and it is called axon hump. The axons conduct electrical signals along the length of the axon mound. These electrical signals are considered as the action capability. [11][12]

Then I will introduce about the artificial equivalents of biological neurons, which are considered as the units in the circumscription. There is an archetypal example about it is concluded in Fig.1.5. It shows that synapses are always modeled by one single number, and we can call it weight. So it is clear that every input is multiplied by the fixed weights, after that, it will be transmitted to the equivalent of the unit body (soma). [11] It means the weighted signals are summed together by simple arithmetic addition. And it intended to provide node activation. It can be shown in Fig.1.5, through that, we can get the so-called threshold logic unit (TLU), and then can use the method by comparing the activation with the threshold; if the activation is larger than the threshold, the cell produces a high value output (normally "1"), which in turn produces zero. In Figure 1.5, the size of the signal is represented by the width of the corresponding arrow, and the weights are shown by the multiplicative symbols in the circle, and their values are proportional to the size of the symbol; only when positive weights are used. TLU is the simplest model of artificial neurons.

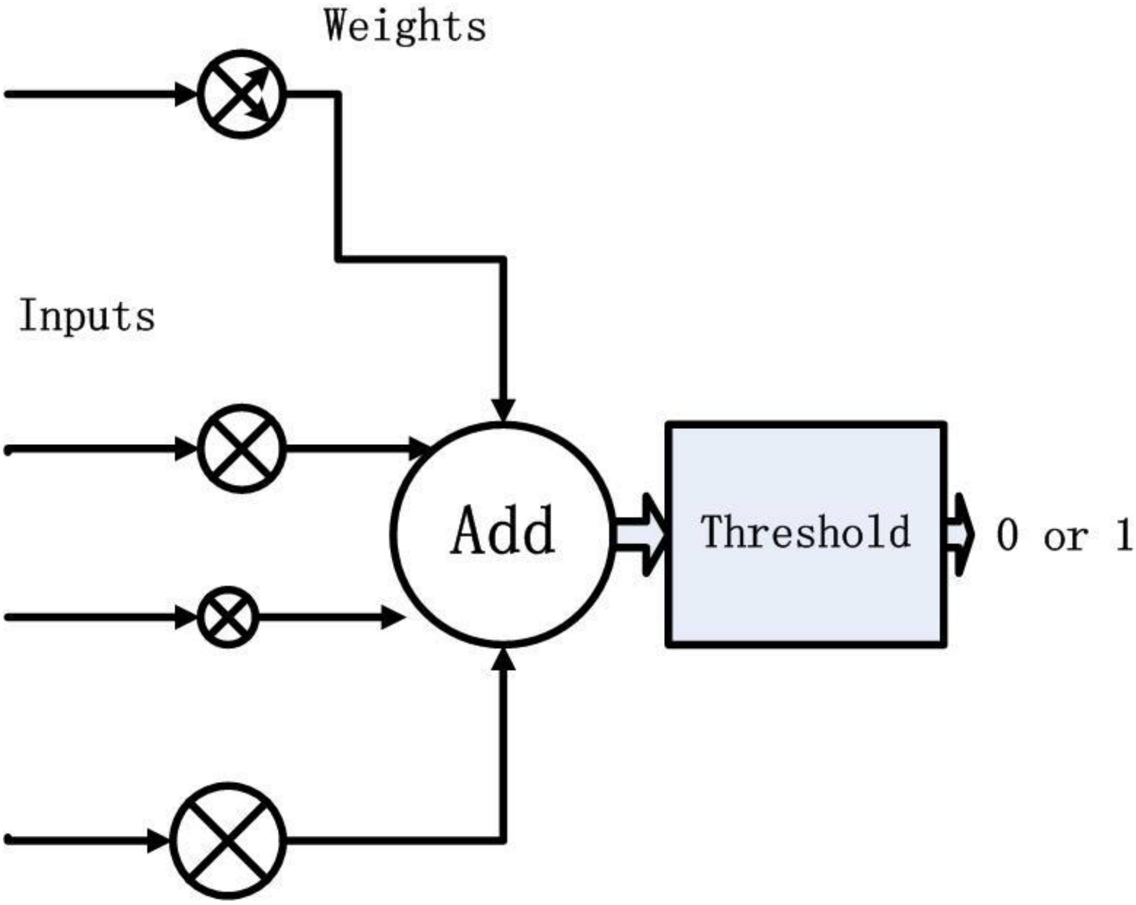


Fig.1.5 Common artificial neuron

And almost all of the artificial neural networks that we know are based on this model. Accordingly, ANNs (artificial neural networks) are not even close to the complexity of the model's brain. However, they have been shown to be easy problems for humans, but for traditional computers, such as image recognition and prediction are so difficult that should be based on past knowledge.

The word "network" is often used to refer to the artificial neuron system. Thus, it can range from a simple single unit to a large number of the collection of units, where each node connects to other node in the network. A type of network may be illustrated as Figure 1.6 shown. Every node here is displayed by just one circle. However, all of the connections maintain the weights. So the nodes are arranged in a hierarchy where each signal radiates from the input, and it passes through two nodes, and then outputs beyond its no longer transformed. The feed-forward configuration is just one of several available, and is typically used to place the input pattern in one of several categories by according to the timing of the output. For instance, if the encoding of both the light and the dark patterns are included in the image of the handwritten letter, the uppermost output layer in the picture will contain 26 nodes instead, and each letter of the alphabet may be used to mark the letter class to which the input character belongs. This can be done by categorizing one output node for each class, and just one of such nodes is triggered whenever a pattern of related classes is provided at the input. [13] There are so many basic structural elements and operations. Back to our definition, we can note the emphases of experiential learning. In real neurons, in some cases, the synaptic strength can be modified so that the behavior of each neuron can be adapted to its particular ascending input. In artificial neurons, this is equivalent to a modification called a weight value. No computer program is available to process the information. The " knowledge" possessed by the network should be stored in its weight, which evolves by adapting itself from some examples to provocative processes. In the training example, it is referred to as supervised learning, one of which is shown in Figure 1.6, where the input pattern is represented as a network whose reply is compared with the target output. The difference which is made by the two output modes can then determine how the weights are transformed. Every specific change recipe builds a learning rule. When the required weight update is performed, another mode is indicated, and after the output compared to the target, a new change will occur. This event array iterates over and over until the network behavior converges. So we can see that the answer to each model is near the target. The process of any sort of pattern presentation and the criteria is used to expire the process. If the network has learned the infrastructure of the problem area, it should correctly classify the invisible patterns and the network is considered to be good. If the network does not have this attribute, and it exceeds the training set classification lookup table. Therefore, good generalization is one of the key characteristics of neural networks.

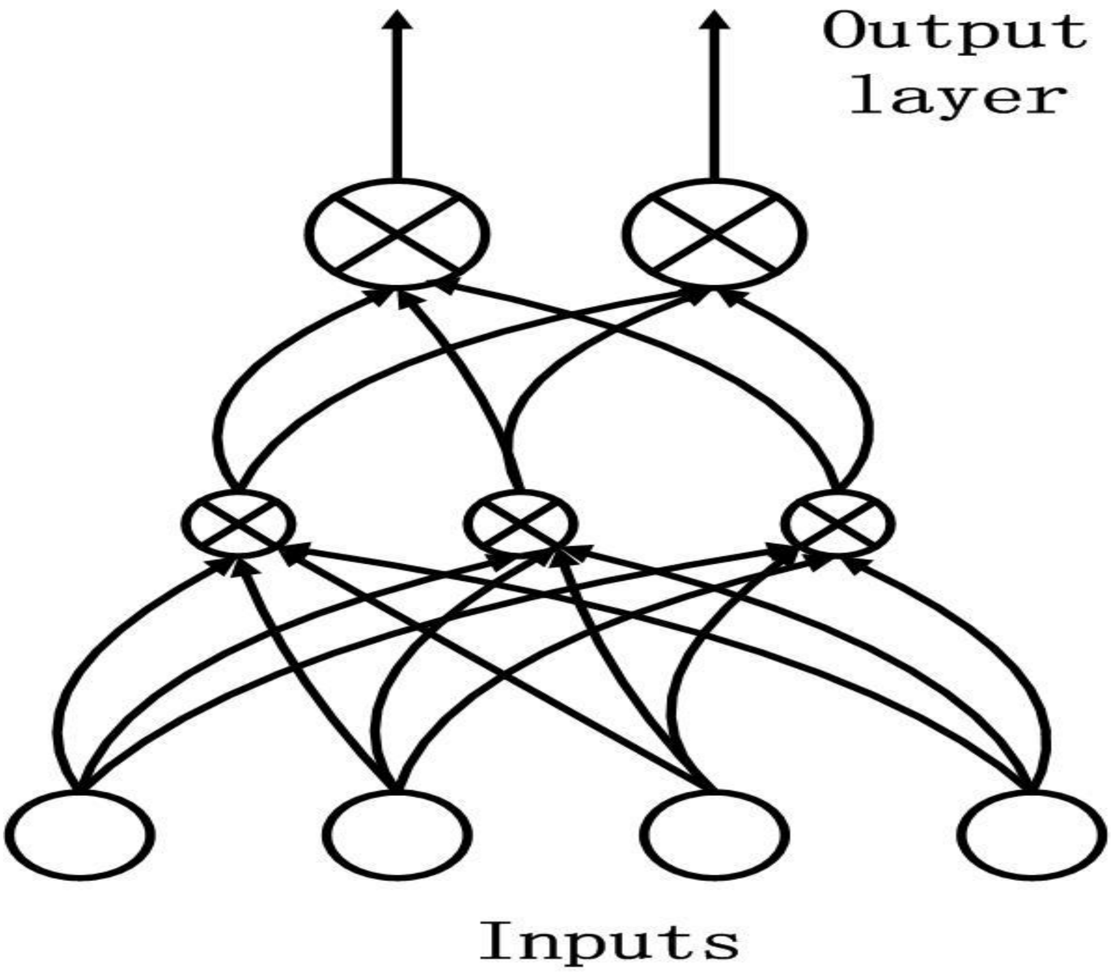


Fig.1.6 Common example of neural network

Artificial neural networks that well known in nowadays appeared after the introduction of the simplified neurons. The concept of simplified neurons is first introduced by McCulloch and Pitts in 1943.[13]

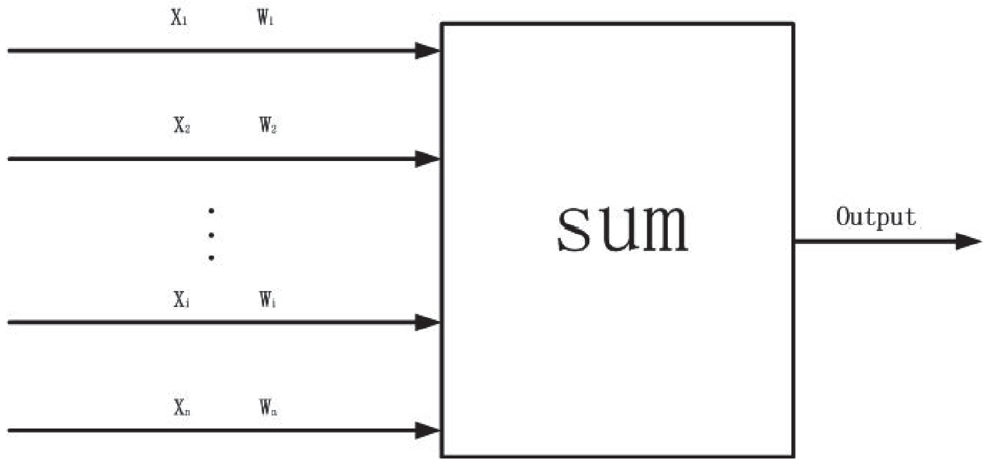


Fig.1.7 McCulloch-Pitts Model.

These simplified neurons are introduced as models of biological neurons, or as conceptual components that can undergo detours of computational tasks. The basic model of neurons is created in the function of biological neurons.

When building a functional model of a biological neuron, there are 3 basic components that should be instated. First, synapses of neurons. they are modeled as weights. And the intensity of the relationship between the input and the neuron is indicated by the value of the weight. Here, the negative weight value forms a known suppressed connection, otherwise, the positive value forms a known excited connection. The next 2 components mimic the actual activity of neuronal cells or somatic cells. The adder adds all of the inputs modeled by their own weights. The activity is treated as a linear combination. Finally, the activation function dominates the amplitude of the neuron's output. The appropriate range for the output is usually between 0 and 1, or between -1 and 1. Mathematically, the process is described in detail in Figure 1.8.

We can see the interval activity of the neuron from this model, and it can be shown as:

$$net_j = \sum_{i=1}^{n} w_{ij} x_i \tag{2.5}$$

Therefore the output of the neuron o_j is to be the outcome of some activation functions on the value of net_j. One often used activation function is sigmoid function:

$$o_j = f(net_j) = \frac{1}{1+e^{-x}} \tag{2.6}$$

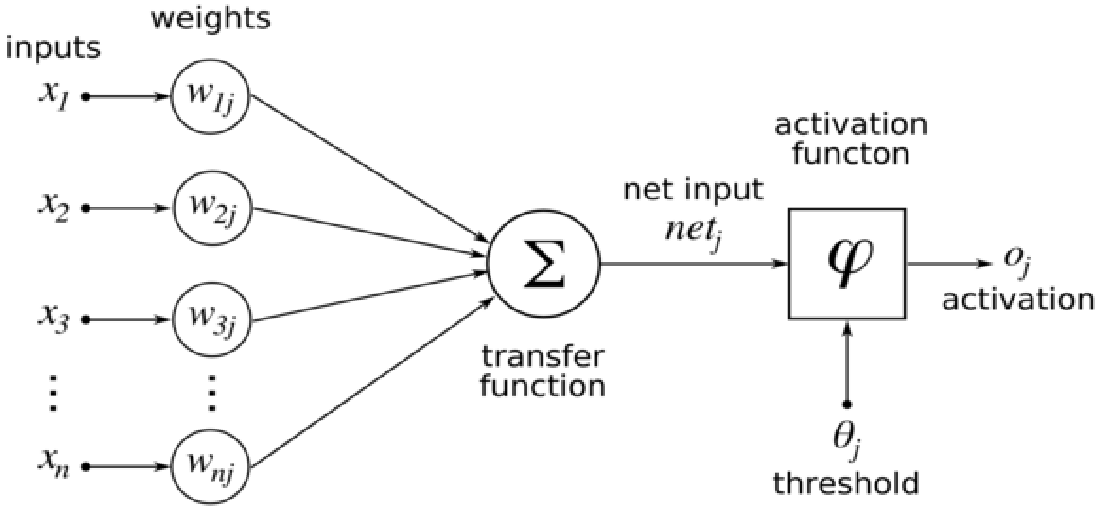


Fig.1.8 Structure of an artificial neuron

1.3.2 History of ANN

In the 1940s, neurophysiologist Warren McCullock and mathematician Walter Pitts wrote an article on how neurons work. For knowing how the neurons work in the brain, they use circuits to model a simple neural network. [14] Then, Donald Habb published his essay "Behavioral Organization," which points out a work that points to the fact that neural pathways are reinforced every time a concept is used that essentially indispensable for human learning in 1949 year. [15] If the two nerves both get angry at the same time, he retorts that the relationship between them is promoted. In the 1950s, with the development of computer more and more advanced methods can simulate a hypothetical neural network. Nathanial Rochester from the IBM research lab firstly proposed a method of simulating a hypothetical neural network. [16] However, his first attempt failed. After that, Bernard Widrow and Marcian Hoff from Stanford developed models instead. And they are called "ADALINE" and "MADALINE". Although the system developed by them is so old that like the air traffic control system, it is still commonly in used. [17]

In 1960s, Widrow and Hoff mature a learning program. It can check the value before weight adjustment (sometimes 0 or 1 in experiment) according to the following rules: Weight change = (pre-weight line value) * (error / (number of inputs)) [18]. The basic idea is that when an active sensor may have a large error, in this case, the weight value may be adjusted to span the network or at least to the neighboring perceptron. If a maintenance error occurs, all errors are distributed to all weights, not the error being cleared. During the same period, a paper was written which suggested that it could not be extended from single-layer neural networks to multilayer neural networks. Furthermore, more and more people use an essentially flawed learning function in the field because it is indistinguishable over the entire line. Due to that, research and funding have declined significantly.

This is in fact related to the fact that before some neural networks succeed, the potential for neural networks leads to an exaggeration, especially considering the actual technology at the time.

In 1972, Kohonen and Anderson gave a similar network on their own. They all use matrix math to describe their ideas separately, but none of them are aware of what they are doing to create an analog ADALINE circuit array. [19] Neurons should activate a set of outputs, not just activate one output. So the first multi-layer network was proposed in 1975, which is an unsupervised network.

In 1980s, more and more interests in the field of neurons was with an innovation. John Hopfield of the California Institute of Technology presented a paper to the National Academy of Sciences. In his paper, the approach was to create more useful machines by using bi-directional lines. However, there is only one way to the connection between neurons. Then Reilly and Cooper used a "hybrid network" with multiple tiers in the same year, each with a different problem-solving strategy. In the same year, there was also a cooperative and competitive neural network of US-Japan cooperation meetings. Japan announced a new fifth-generation effort neural network involving artificial intelligence, the United States file to worry about. The United States is afraid of being abandoned in this leadership. The first generation uses switches and wires, the second uses transistors, and then the third state uses solid-state techniques. Therefore, there are more foundations and more research in this field. [20]

In 1986, with the development of multilayer neural networks in the news, the question of how to extend the Widrow-Hoff rule to multiple layers was mentioned. 3 independent research teams have grabbed the analogous idea, now known as the Back Propagation Network, one of which involves David Rumelhart who was a former member of Stanford's psychology department because it misidentified patterns throughout the network distribution. The hybrid network uses only two layers, and these back propagation networks use a lot. The result of its turn is that the back-propagation network is "slow learner" because it needs to be able to learn thousands of layers.

Nowadays, neural networks are used in some areas as we all know, in my thesis, I will describe some applications later. The basic idea of neural network properties is that if it works in nature, it must be able to work on a computer. Therefore, the future of neural networks is related to the development of hardware.

1.3.3 Why study neural networks

The question is applicable here because, according to one's intentions, the study of connectivism can substitute for different points of view. Neural networks are often used for statistical analysis and data modeling where their roles become aware of alternative standard non-linear regression or beam analysis techniques. Therefore, they are often used for classification or prediction problems that may arise.

Some examples, such as image and speech recognition, text character recognition, and human attitudes, including medical diagnostics, petroleum geology, and financial market indicators are mentioned. This type of problems is also in the realm of classical artificial intelligence (AI), therefore, engineers and computer scientists have noticed that neural networks provide parallel distributed computing styles that, in turn, provide an alternative to machine intelligence.

By a concise explanation of this term, the correspondence refers to the fact that each node is assumed to operate on one itself and operate concurrently with the other nodes, and the " knowledge " in the network is distributed rather than concentrated as a conventional computer in several memory locations.

Practitioners in this field do not care about biorealism and are often due to the ease of implementation of the solution in digital hardware or the efficiency and accuracy of certain techniques. Haykin (1994) from the engineering point of view gave a variety of neural network technology for a comprehensive survey. [21]

Neuroscientists and psychologists are interested in networks. Because the networks are like computational models of animal brains, and it is developed by extracting those qualities that are necessary for the information processing of true neural tissue. However, only time will tell us that by making use of how real neuronal interconnection as a local "circuit" of knowledge, functional modeling of the brain has made great progress. A good introduction to this procedure for calculating neuroscience is by Churchland & Sejnowski (1992).

There should be mentioned in the last that more and more physicists and mathematicians have traction to networks of interest in nonlinear dynamical systems, statistical mechanics and automata theory. The work of applied mathematicians is to use the tools previously used in other fields of science to discover and systematize the properties of new systems. For example, there is a strong link between some type of mesh and a magnetic system called rotating glasses. The complete mathematical device used to explore these links was developed by Amit (1989) (along with a series of concise summaries).

Every group is appealing different questions. Neuroscientists intend to know how the animal brain works, while engineers and computer scientists who intend to understand the basic properties of networks as complex systems are attend to build intelligent machines and mathematicians. Another (or perhaps the largest) group of people can be found in various industrial and commercial fields, they are giving their interests to the simulation and analysis of naturally occurring in the workplace in the use of neural network, however, it is difficult to understand the data set. Therefore, it is very important to make sure the author's viewpoint when reading literature. However, their common focus in neural networks may be the basis for close cooperation. For instance, biologists can effectively learn what scientists from the computer calculation, allowing the animal to solve specific problems, solutions and engineers can use natural design, so that they can be applied to "reverse engineering" behavior.

ANNs (artificial neural networks) can be thought of as a simplified model of neural networks. And they naturally exist in animals' brains. Through a biological point of view, the basic requirement of a neural network is that it should try to capture the basic information-processing features that we consider to be the corresponding "real" networks. For engineers, this communication is less important after comparing, and the network is provided by an alternative form of parallel computing. Form the experience, it may be more appropriate to solve the current task. The simplest artificial neurons are threshold logic cells, or we can call them TLUs. The basic operation of the TLUs is to perform a weighted sum of its inputs. And we can get the output "1" when the sum exceeds the threshold, otherwise, the output will be "0" instead. TLUs should simulate the basic "integration and excitation" mechanism of real neurons.

1.3.4 Problems of Neural Network

Although it has been proven that neural network is an effective method to solve combinatorial optimization problems, neural network still has some problems preventing its development. [22] The main problems are as follows:

- Parameters selection problem.

It is difficult to select an effective parameters to initial neural networks when using them to solve combinatorial optimization problems. Such as illustration of Wilson and Pawley, They could not find that the parameters of the Hopfield model need to be changed as the size is enlarged, because no combination of parameter values is always found that produces a valid solution of executive efficient problem

Some neural networks are able to find the optimal solutions of optimization problems, but they should spend long executive time and cost large numbers of resources. Although many researchers have proposed some improvement methods to increase neural networks' efficiency, but the effect is not clear.

- Local minimum problem

As a proximate algorithm, neural network is easy to get convergence to a near optimal solution. These solutions lead to local minimization problems, which is one of the major drawbacks of Hopfield neural networks. Because the energy function consists of several terms, there are many local minima, and there is a trade-off between these terms. When at least one of the constraint penalty items is non-zero, an infeasible solution to the problem arises.

1.4 Statistics and Neural Networks

In several instances, it has already been noticed that there is a similarity between the training algorithm of the network and some of the techniques in statistics or data analysis. For instance, we can refer to back propagation as non-linear regression and competitive learning to certain types of clustering analysis. Are all the neural networks just a familiar technique for redo situations? We claim that although there is a similarity between the "classic" approach and the network, the latter does retain the novelty and utility beyond the established method.

Consider the MLP definition again. At the calculation level, this may be the model is considered to be a non-linear regression with parameterization by the weights. But in 1994, Cheng & Titterington pointed out that it is not consistent with any previously developed regression model, and is not a priori, but network-based. [23] Thus, network paradigms have the potential to stimulating novel computational methods in a bottom-up fashion; it isn't just a given implementation of a computational strategy, and indeed can support its own merits.

Suppose that we find a network in the biological environment, which may have the same structural properties. For those we should study its artificial disguise. If we have a network signal-level implementation that gives it biological plausibility, there are further benefits.

The feed-forward network is proposed to select the best function which can perfectly fit a set of input-output data. The changes related to the network weights permit tuning of network functions so that can detect the best configuration. There are 2 complementary motivations that can decide our view of the best means in the case. On the one hand, we expect the network can map the known inputs as accurately as possible, so that we can be aware of the outputs. On the other hand, the network must be able to generalize the unknown inputs which will be compared to known inputs, and through that, it will lead the resulting output to be the interpolation of the learned values.

From the Fig 1.9, we can see that the problem from another point of view. The points in Fig 1.9 represent the training set. A function that can map a known input to a known output are being searched. If using a linear approximation, as shown, the residual is not too large, and a new unknown value of the input x can be mapped to the regression line. When Figure 1.10

shows that there is another approximation of the function, it is using linear splines which can reproduce the training set without error. While the training set is a consist of experimental points, there is usually some noise in the data. Therefore, the linear approximation in Figure 1.9 may be a better alternative to the exact fit of the training data.

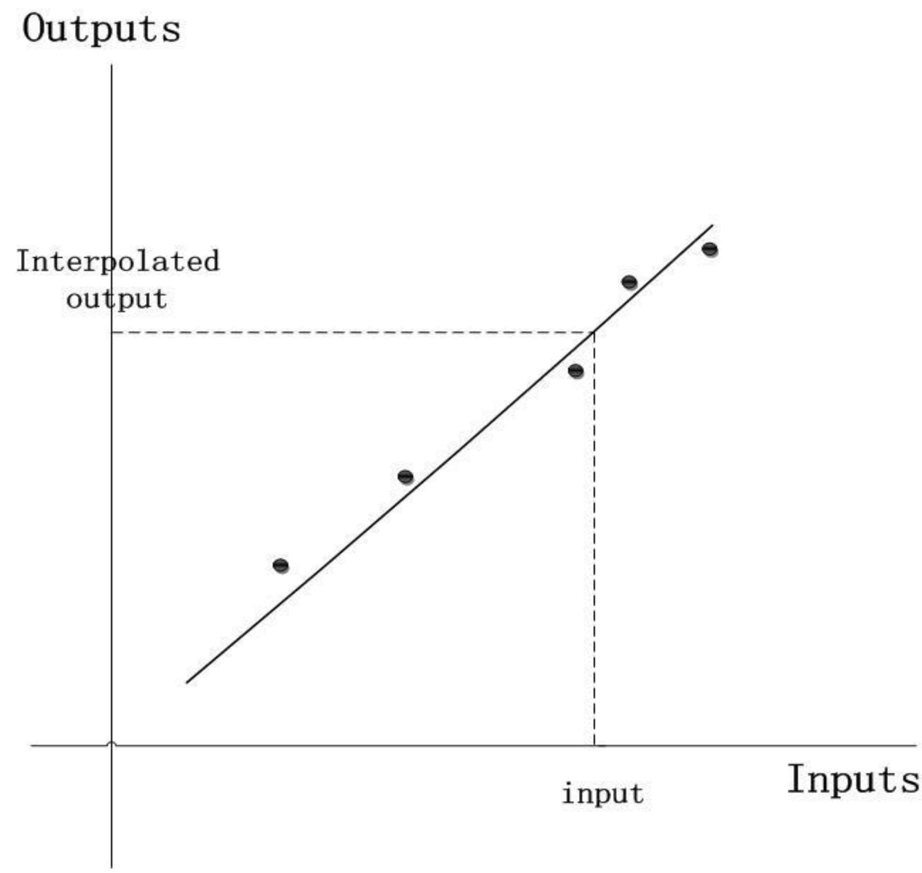


Fig.1.9 Linear approximation of the training set

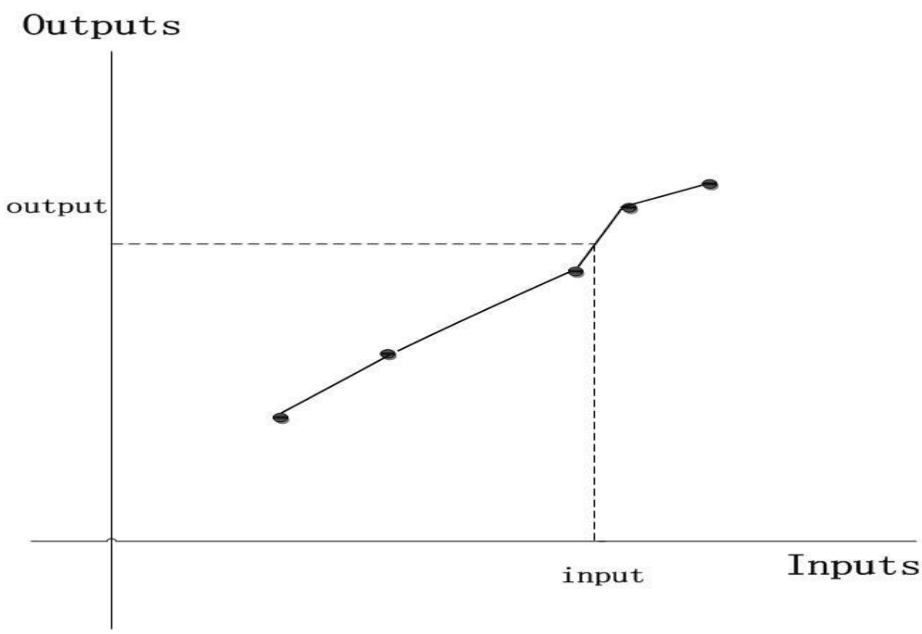


Fig.1.10 Approximation of the training set with linear splines

Here is no general way to decide the optimal number of parameters of the networks. It is totally decided by the structure of the current issues. The best results are obtained when the network topology is selected, and it can take into account to know interrelationship between the inputs and the outputs. As the examples shown in the above, the linear approximation would be the best if the theoretical analysis suggests to speculate the linear correspondence between the input and the output, although the multi-linear approximation had less training error. Statisticians have studied this function approximation for a given training set in the field of both linear and non-linear regression. In a sense, the back-propagation algorithm is only a numerical method for statistical approximation. Analyzing the linearity can improve our understanding of the connection.

A linear correlator is a computing unit that only adds its weighted input. We can also think of them as part of a nonlinear element. If it has a linear correlator for the weight vectors (w_1, w_2, ..., w_n) of the n-dimensional input (x_1, x_2, ..., x_n), the output can be expressed as $y = w_1x_1 + \ldots + w_nx_n$. As Figure 1.8 shown, the output function of a linear correlator, and the correlaor has two inputs. The learning problem is the output from the input vector in the reproduction training set. The points will correspond to the training set after training. The parameters of the hyperplane should be chosen to minimize the error. The back-propagation algorithm can be used to find them out.

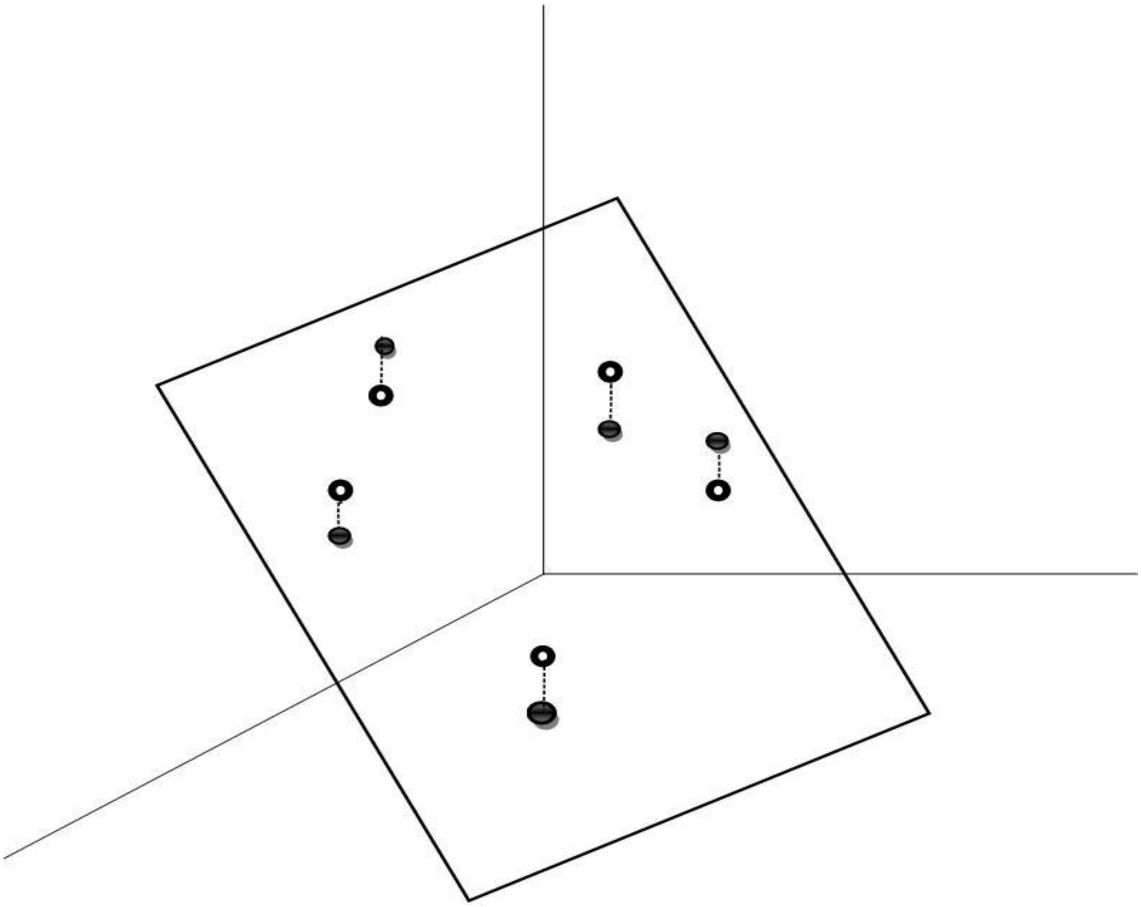


Fig.1.11 Learning problem for a linear associator

This paper is organized as follows: In section 2, the models that the research proposed are described. They can be divided into three types: statistical models, neural network models and combination model. Section 3 and section 4 separately introduces the experiment of time-series forecasting for house price index of China and tourism economy. Section 5 provides concluding remarks.

2. Modeling

The time-series model explains variables about its own past and stochastic perturbation terms. Time series models have been widely used for a lot of areas' forecasts over the last four decades. In this section, statistical models, Artificial Neural Network are described as follow.

Traditionally, time-series prediction has been primarily performed using statistical models.

2.1 Statistical models

In recent years, statistics has been widely applied to the time-series forecasting. Among the statistical methods, the most popular methods are the Naïve model, the exponential smoothing model (ES, and the autoregressive integrated moving averages model (ARIMA). Among them, the most advanced forecasting model is Autoregressive integrated moving-average model (ARIMA) which has been successfully inspected in many theoretical and practical applications. If the linear models have a good approximation to the underlying data generation process, they can be regarded as the preferred models.

2.1.1 The naïve model

The naïve model is the most cost-effective forecasting model, and provides that a benchmark can compare more complex models. This prediction method only applies to time-series data which uses the naive model to produce a prediction equal to the last observation. This approach is very efficient for economic and financial time series, which often have patterns that are difficult to predict reliably and accurately. Seasonal naïve methods may be more appropriate if the time series is considered to be seasonal, where the forecast is equal to the value of the previous quarter. The naive model may also use drift, which will use the last observation plus the average change from the first observation to the last observation. The naïve forecasting model can be simply stated that the forecast value for the period F_t , it is the same to the actual value of the last period available value (X_{t-1}). The equation can be shown as follow:

$$F_t = X_{t-1} \tag{2.1}$$

Where F is the predict value, X is the truevalue, t is some time period.

Eq.2.1 can be considered as a random walk model in which the assumed trend and the turning points that can not be forecasted and the prediction can be seen as horizontal line extrapolation.

2.1.2 The ES model

The ES model (exponential smoothing) is the most commonly used time-series prediction method. It was based on the development of mobile averaging technology. It predicts the effect of the closest actual value on the predicted value based on the current actual value and the currently predicted next value but does not require a large amount of past values, it can be shown as:

$$F_{t+1} = aX_t + (1 - a)F_t \tag{2.2}$$

Where F_{t+1} is the prediction value of time interval t+1; a is the smoothing constant (when $0 < a < 1$). X_t is the actual value when in the time t, F_t is the prediction value of time interval t. From Eq.2.2, We can find that a can be regarded as the weight of the past history. The larger of a value, the less weight the past history has. The history is related to the last actual value of the parameter. This method is referred to as "exponential" because the predicted value is a discrete convolution of the observation sequence with an exponential curve with a time constant $1/(1-a)$. In turn, if the value of X_t becomes stable, the error $(X_t - F_t)$ decays exponentially.

2.1.3 The ARIMA model

ARIMA is the most welcomed model of the linear model. It is used to predict time-series. It has been a large success not only in the academic research but also in the industrial and economic applications. A common ARIMA model is ordered by (p, d, q), and it can be shown as:

$$\phi(B)\ \nabla^{d} x_t = \theta(B)\ \varepsilon_t \tag{2.3}$$

Where x_t and ε_t is the number observations and the random error terms at time t respectively. B is defined by $Bx_t = x_{t-1}$, here it is concerned to ∇, when $\nabla = 1 - B$, $\nabla^{d} = (1-B)^{d}$, and B can be considered as a backward shift operator. d represents the order of differencing. Here $\phi(B)$ and $\theta(B)$ separately represent autoregressive (AR will be used hereinafter) and moving averages (MA will be used hereinafter) operators of orders p and q respectively, and they can be defined as:

$$\begin{aligned} \phi(B) &= 1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p \\ \theta(B) &= 1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q \end{aligned} \tag{2.4}$$

When $\phi_1, \phi_2 \ldots \text{-} \phi_p$ are the autoregressive coefficients, while $\theta_1, \theta_2 \ldots \theta_q$ are the moving average coefficients.

When the ARIMA model is fitted to the original data, the ARIMA model should go through the following 4 steps.

(I). Identify the ARIMA (p,q,d) structure.

(II). Estimate the unknown parameters.

(III). Goodness - of - fit test for the estimated residual.

(IV). Predicting future results based on known data.

The ε_t shoud be independent and it can be considered as a positive random variable with mean = 0 and a constant variance=σ^2. The roots of $\emptyset_p(x_t) = 0$ and $\theta_q(x_t) = 0$ should all be outside the unit circle. Box and Jenkins (1976) suggested that the ARIMA model should use at least 50 or preferably 100 observations.

If the data has significant seasonal changes periodically. We can use the SARIMA model to eliminate the effects of seasonal cycles.

2.2 Neural Network models

In this section, I will introduce five models of neural networks as follow.

2.2.1 Hopfield Neural Network

In popular terms, "remembering" things involves associating thoughts or ideas with sensory cues. For instance, somebody may notice the name of a celebrity, thus we immediately recall celebrity television or newspaper articles. Or we may see pictures of the places we visit, which recall the memory of the people we meet and the experiences we enjoyed at the time. (Smells) can also be recalled and considered to be particularly effective, so we will consider a more trivial example, including all aspects. Consider the left side of Figure 2.1. This should represent the binary version of the letter "T". The open and filled circular symbol indicates that the pattern in the center of the 0 and the 1, then the graph is the same "T", but the lower part is replaced by a noise with a probability of 0.5. We can imagine that the upper part of the letter is provided as a cue, and the lower part must be remembered. Get the correct pattern from the original "T"; by adding 20% noise, each pixel is inverted with a probability of 0.2. In this case, we assume that the entire memory is available, but in the form of an incomplete adjustment, the task is to "remember" the original letter in its undamaged state. [24]

It can be compared to our "fuzzy" or inaccurate memories of some scenes, names and sequences of events. It also can be stitched together after recalling some efforts.

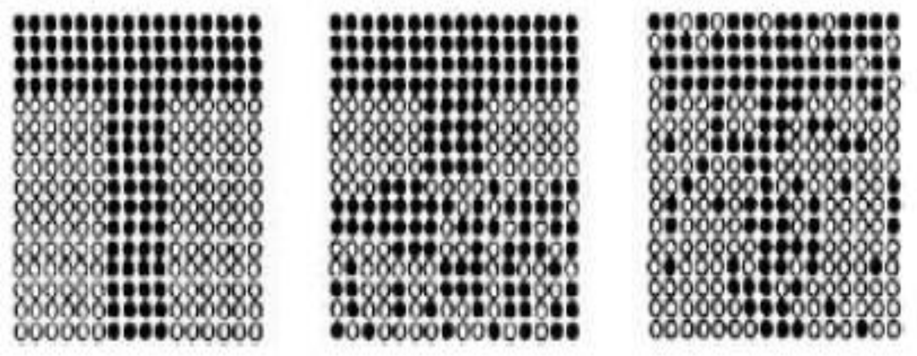


Fig.2.1 Associative recall with binarized letter images

The general examples herein can be concluded as follows. There are some basic stored data sets that are sorted and interrelated to some extent; the data constitutes a stored pattern of memories. In the above example of human memory, it is a set of items which are associated with celebrities or places we visit. It is part of some alphabet (pixels), the arrangement of which is determined by the stereotyped version of the letter by the case of character recognition. Alternatively, an incomplete version of the stored memory may be provided to us, which must be associated with the actual undamaged mode. Note that it is not important which part of the pattern is used as a hint; because the entire model is always restored.

Sometimes, we can search for any of these discrete items by selecting the correct option from the menu and entering the complete project. Suppose now we only have code to record the fragment "ion, Mar" in "Vision, Marr D". The database cannot use this snippet or even start the search. We do not know whether it belongs to the author or maybe to the title, even if we do so, we may also find "leave" at the beginning of the title or author. The input to the regular database must be very specific and complete after it working like this.

Considering a feed-forward network where the output target of it is the same as its input. This type of network can be considered as associative memory, since an incomplete copy of the training sets should result in a true vector at the output from which it is derived. The network is the first to be used for memory storage and its mathematical analysis can be found in 1982 which is proposed by Kohonen. However, John Hopfield insisted that there is a potentially more powerful network type for associating memory popular, and it is different from the above-described network types because the network has a feedback loop in its connection path. The relationship between the two types of association networks is discussed below. The Hopfield network is in fact an example of a more general class of dynamic physical systems that can be considered to instantiate a ”memory”, as a stable state associated with a minimum value of an appropriately defined system energy. Therefore, we now talk about the description of these systems.

In the early 1980s, two scientific papers were published by Hopfield, they aroused great interests and became the starting point for a new period of neural networks and it continues until today.

Hopfield shows that the physical system model can be great approach of solving computational problems.

Hopfield network is fed by the feedback of the whole network, in the form of the network by the neurons are connected with each other, that is, each neuron will own output through the connection right to other neurons, while each neuron accepts information came from other neurons.

In the acyclic network, the information is added from the input end of the network, through the network processing step by step, the final output from the output, this process does not exist signal feedback.

In the loop network, the network to receive a signal, it makes the signal in the network through repeated iterative processing, until the change stops, or changes in the amplitude is small enough, the network can be given at this time the corresponding output can be regarded as its output. Obviously, for a given input, the network output is constantly changing from its feedback signal caused. The processing of the input signal by the network is a gradual restoration and strengthening process. [25]

We have to attempt to apply the above concepts to the construction of neural networks. And it is capable of performing associative recall. Considering there is a network of 3 TLU nodes, as Fig.2.2shown. Every node is connected to each other, and the connection strength or weight is symmetric because the weight from node i to node j is the same as the weight from node j to node i. it can be presented as $w_{ij} = w_{ji}$ and w_{ii}=0 for all i, j. In addition, the threshold is assumed to be “0”. Since the signal may flow back from one node to itself via another node. We can say there must be feedback in the network, or we can say it is iterative. Because nodes can be used repeatedly to process information. This is in contrast to the only feed-forward network used to date.

This type of network and its energy-based analysis was elegantly talked about by John Hopfield in 1980s, thus his name is often related with this type of network. However, there is some other model that is very close to the "Hopfield model". It was proposed by Little before 1974, but there is no emphasis on energy-based descriptions here.

Little has also been widely used in quantum mechanics, which may make his work more inconvenient for non-physical readers.

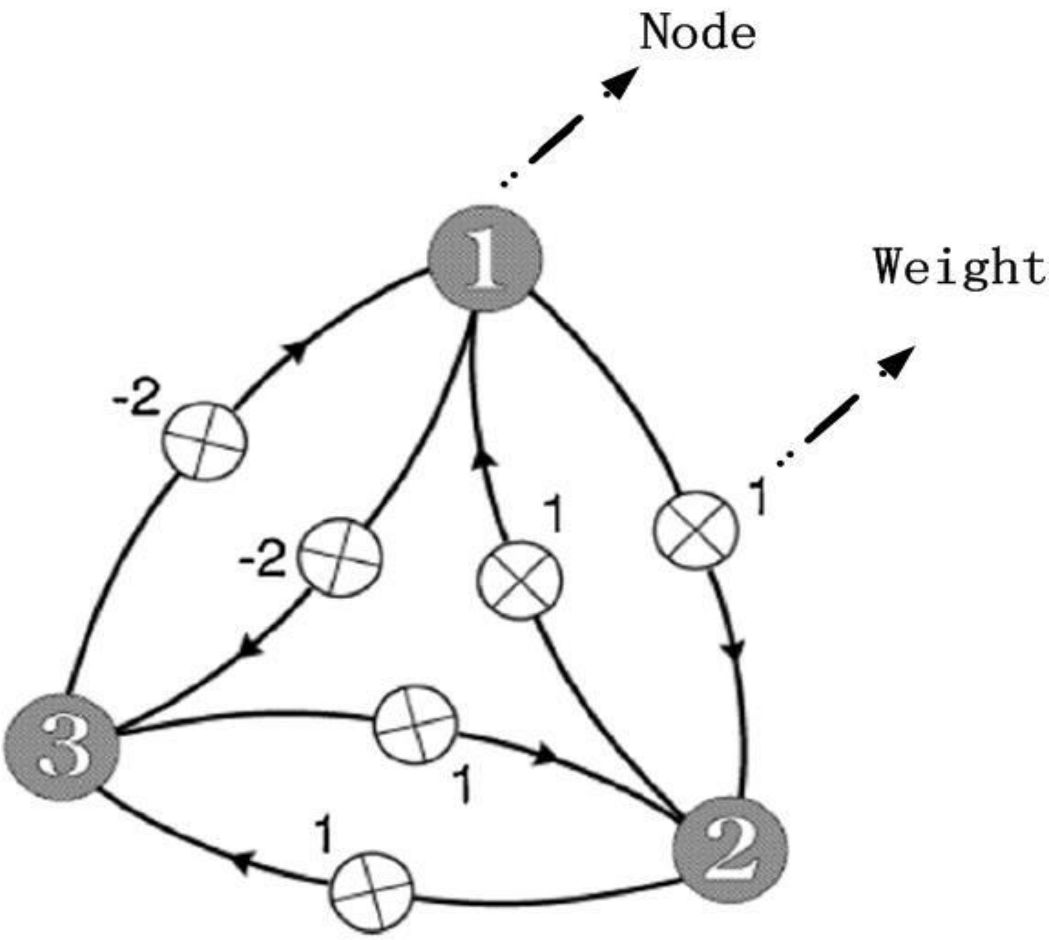


Fig.2.2 Three-node Hopfield net

In the n-dimensional data space, the classical pattern should have n-binary components, it can be shown as 1 or -1; this means that each classical pattern corresponds to the angle of the cube in the n-dimensional space. Then the network is used to classify the distortion patterns into these classes. While the distortion mode is used to the network, it must be associated with another mode. As the network is working properly, then the association pattern is one of the class patterns. Sometimes pseudo-minimum values may also occur.

Hopfield networks are always referred to as associative networks. The reason is that they often associate a class pattern with each input pattern.

Hopfield networks have 2 types of neural networks: one is the continuous-time versions, and the other is the discrete-time versions. The two network types have a weight matrix W defined as:

$$\mathrm{W} = \frac{1}{n}\sum_{i=1}^{D} \vartheta_i^T \vartheta_i \tag{2.7}$$

Where D is the number of class patterns ($\vartheta_1, \vartheta_2, \dots, \vartheta_D$), and n is the number of components.

Hopfield networks of discrete-time versions have the following dynamics:

$$\mathrm{x(t+1) = Sign[Wx(t)]} \tag{2.8}$$

For a discrete-time Hopfield network, x is given by the follow equation:

$$\mathrm{E(x)} = -\mathrm{xW}x^T \tag{2.9}$$

We can realize that if an initial state vector x(0), x(t) is given as in Eq.2.8, it will converge to a value which has minimum energy. Through that, the minimum of Eq.2.9 constitutes the possible convergence points of the Hopfield network, and ideally, these values are the same as the class pattern ($\vartheta_1, \vartheta_2, \dots, \vartheta_D$). Therefore, it is guaranteed that the Hopfield network will converge to a certain mode, however there is no guarantee that it will converge to the correct mode. However, this is only a scaling issue. Add large enough constants to the energy expression to make it positive.

Continuous Hopfield networks are stated by the following differential equations:

$$\frac{dx(t)}{dt} = -x(t) + W\sigma[x(t)] \tag{2.10}$$

Here x (t) presents the state vector of the network, when W is a parameter weight. Sigma is the nonlinearity acting on the state x (t). The weight W is defined in Equation 2.7. The differential equation (Eq.2.10) is solved using an Eulerian modeling.

For a continuous-time Hopfield network, it will be defined by the parameters given in Eq.2.7:

$$\mathrm{E(x)} = -\frac{1}{2}xWx^T + \sum_{i=1}^{m}\int_0^{x_i}\sigma^{-1}(t)dt \qquad (2.11)$$

For a discrete-time network, it can be shown that the state vector x (t) in Equation 2.10 converges to the local energy minimum for a given initial state vector x (0). Thus, the minimum of Eq. 2.11 constitutes a possible convergence point of the Hopfield network, and ideally these minimum values coincide with the class patterns ($\vartheta_1, \vartheta_2, \dots, \vartheta_D$). However, there is no guarantee that the minima will match with this set of class patterns.

2.2.2 Self-Organizing Map

In this section, we explore the possibility that the network can discover clusters of similar patterns in unsupervised data. It means, different from the MLP, the target information is not provided in the training set. A typical data set to be processed is schematically illustrated as Figure 2.3 shown. The point appears to fall naturally in 3 clusters, two of them are smaller, they are more closely bound to the left, and the left larger one is more closely bound to the right. Pay attention that there is no mention of class labels. Because of that it is known as experiencing self-organization and unsupervised learning by assigning nodes to nodes in some way.

The key technologies used in training networks in this way involve nodes that are most responsive to any pattern. One approach is to simply search for the largest activity in the network. When shifts the responsibility of this process to some sort of oversight mechanism that is not part of the network. Another approach is to provide the network with additional resources. It allows this search to instead of the network itself. This distinction will be clearer in the future, however we are here to introduce it about its motivate in the next section. And we will also deal with the search mechanism within the network.[26]

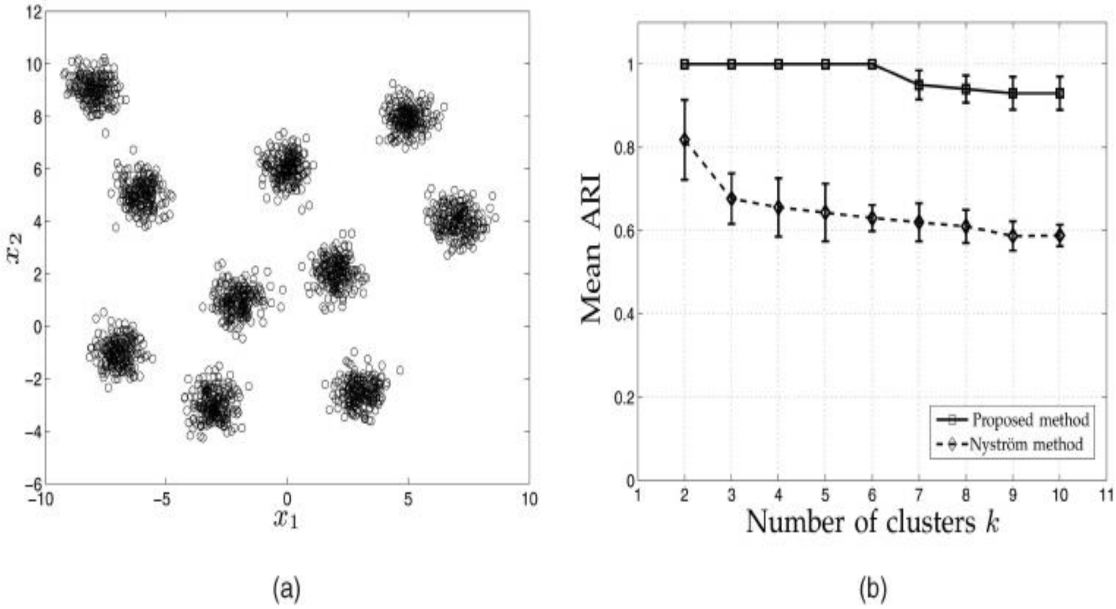


Fig.2.3 Clusters in a training set

Self-organizing mapping (SOM) is one type of artificial neural networks which uses unsupervised learning to train the produce of a low-dimensional (almost two-dimensional) discrete. The representation of the input space of a training sample can be called as a map. This mapping is designed to preserve the topological properties of the input space. [27]

Such as most artificial neural networks, SOM can be in 2 modes: one is training, and the other one is mapping. Training uses the input example to build the map. This is a competitive process, also known as vector quantization. Automatically categorize the mapping of new input vectors.

The components that the SOM contains are called nodes or neurons. Weight vectors of the same size as each node are used as input data vectors and locations in the map space. The usual arrangement of nodes is the regular spacing between hexagonal or rectangular meshes. The SOM description maps from a high-dimensional input space to a low-dimensional space map. The process from the data space vector map is to find the nearest node to the weight vectors of the vector data distribution and the space node coordinates to the vector. While this type of network architecture is generally considered to be associated with a feed-forward network in which nodes are visual connections, and this type of architecture is fundamentally different from the arrangement and motivation.

The SOM's basic idea is simple and effective. The SOM is defined from the high-dimensional input data space to the rules of the two-dimensional mapped neuron array. Each neuron of map is related to an n-dimensional vectors $m_i = [m_{i1}, m_{i2}, \ldots, m_{in}]^T$, here n is considered as the dimension of the input vector. Neurons of the map are connected to neighboring neurons through neighborhood relationships, which indicate the topology or the structure of the map. The most common topologies used are rectangles and hexagons.

The network architecture also includes a set of fully connected to a specific input layer, but now there is no horizontal connection. From the analysis of the directional map in the previous section, it is clear that the principle used to form the mapping relation is that the training should be performed on the extended area network concentrating on the largest active node. What is needed is a conceptual network that defines "communities". This solves the self-organizing layer by the spatial relationship between the internal nodes, as shown in Figure 2.4. And the above two two-dimensional array rectangles and hexagonal grid forms show three community plans. In all cases, three communities separated by shaded cells from nodes 1, 2, 3. Thus, the linear, rectangular, and hexagonal arrays are in their distance-2 neighborhoods (including the central node) have 5, 25, and 19 nodes, respectively. Although the three-dimensional array of nodes is conceivable, due to their complexity, they are often not used in practice.

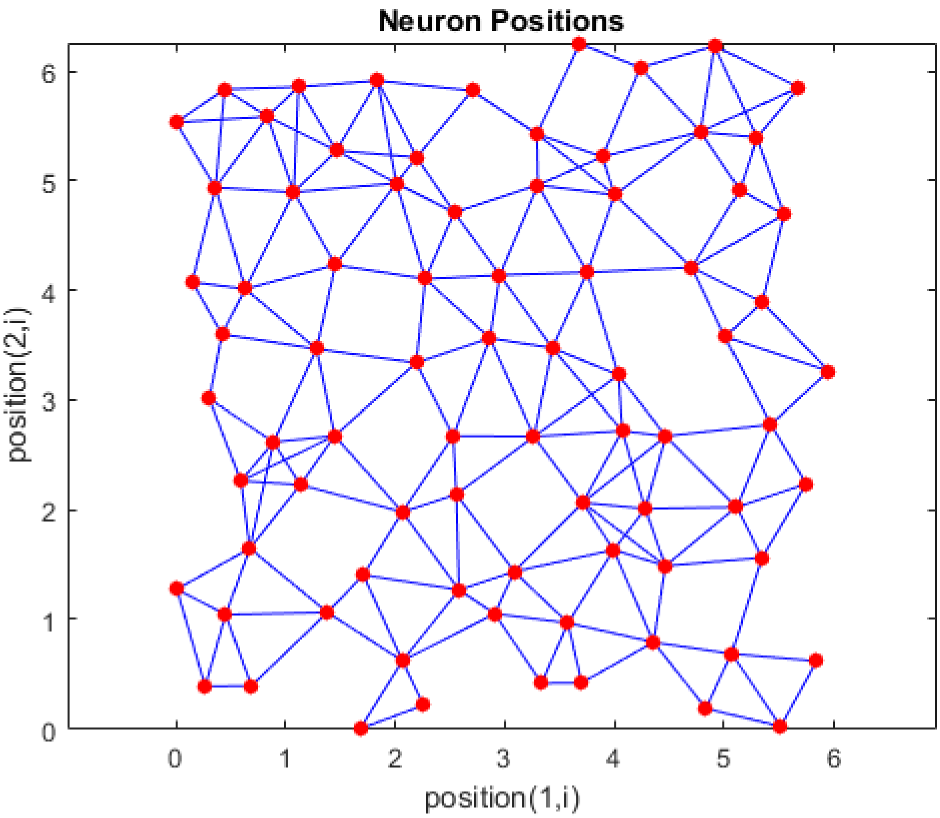


Fig.2.4 Neighborhood schemes for SOMs

In the algorithm of the basic SOM, the number of topologies and neurons is fixed from the beginning. The size of the mapping is determined by the neurons' number which can also affect the accuracy and generalization of SOM.

When in the time of the training phase, the SOM forms an elastic mesh, which is folded onto the "cloud", by the input data. The algorithm controls the network and attempts to approximate the density data. Reference vector in the codebook floats to high-density region of the input data. However, only a few codebook vectors are located in areas where the input data is sparse.

The basic learning process of SOM can be described as follow:

a. A sample vector x is randomly extracted from the input data set, and its similarity (distance) to the codebook vector is calculated by using, for example, a vector. Common Euclidean distance measurement:

$$\|x - m_c\| = min_i\{\|x - m_i\|\} \tag{2.12}$$

b. As learning takes place and the new input vector is given to the map, the learning rate is gradually reduced to “0” according to the specified learning rate function type. As the learning speed, the radius of neighborhood decreases.

The update rule of the reference vector of unit i is as follow:

$$m_i(t+1) = \begin{cases} m_i(t) + \alpha(t)[x(t) - m_i(t)]\,, i \in N_c(t) \\ m_i(t)\,, i \notin N_c(t) \end{cases} \tag{2.13}$$

c. The step a and b together show the number of training steps. It must be fixed before training the SOM since the convergence rate and the learning rate in the neighboring functions are calculated accordingly. SOM, since the convergence rate and the learning rate in the neighboring functions are calculated accordingly.

The horizontal center around the external junction in the layer of neurons, make active contour by contrast enhancement layer. So, setting up have the strongest external input node under the weak driving node price became very active. Use proper transverse weights, which can lead to extreme, the winners have all dynamics, which has the largest external input node to achieve the maximum effective strength, and minimise the activities of all other nodes. This mechanism is used in the competitive learning, which the weight vector of node through them and clustering center alignment and is associated with pattern clustering.

X for any given model, the weight vector w with x alignment node is adaptation, makes the w to move closer to x. Determine the degree of weight design aimed at by the node activity, because the dot product with the two vectors is proportional to the wx (in two) under the normalization scheme of vector set. Therefore, the most active node is its weight vector should be adapt to the competition and you can use for all (winner) dynamically to find the node. Ideally, each cluster by at least one node (given maximum response) encoding. This type of equivalent network, called unsupervised learning and experience the self-organizing process because there are no known target output.

In the topography map, not only response to the cluster, and they are arranged in the network, makes the adjacent nodes coding "close" cluster for each other. The closer to the concept here is based on the model space with basic dimensions of each vector. It can find the singularity or rupture. Cortex of topography map widely exists in animals and their sensory and motor information coding. They can be used under the self-organization in the competitive learning to use the similar development in artificial network learning rules, but now not only training "winning" node, and training in the community around the node. In order to ensure good resolution pattern space good sort of fine details of image, must be conducted with the training and reduce the neighborhood size and learning rate. In this way after training self-organizing mapping (SOM), if known linear vector quantization (LVQ), you can use linear vector quantization (LVQ) to improve its boundary. [28][29]

The purpose of biometric features in the set figure can be related to keep "line length" needs. SOM algorithm shows that the development process of the brain may, characteristic figure allowed in the training of the training set in space visual basic relations. SOM can consider to reduce the size of the input space; Another attempt to do linear statistical method is principal component analysis (PCA), it is shown that a neural network implementation is contained.

2.2.3 Maximum Neural Network

In 1985, Hopfield and Tank first introduced the neural network approach to the travel salesman problem, since then it has been widely used to solve combinatorial optimization problems. In order to avoid local minimum convergence and furthermore, in order to improve the solution quality, the largest neuron model has been successfully proposed by Takefuji et al., Lee et al. in 1992. It deals with a class of NP-complete optimization problems, which is difficult for a neural network to solve. They show the effectiveness of the maximum neuron model with 2 NP-complete optimization problems, namely, module orientation problems and objective functions, such as bus length and sub-graph size, and no constraints are imposed. [30]

The maximum neural network (MNN) consists of M clusters of N neurons. The total number of neurons is M*. One and only one neuron among N neurons with the maximum input per cluster always has nonzero output.[31]

The input or output function of the j-th neuron in cluster i is given by:

$$V_{ij} = \begin{cases} 1\ ;\ if V_{ij} = max_{k=1,2,\ldots m\{U_{ik}\}} \\ 0\ ;\ \ otherwise \end{cases} \tag{2.14}$$

While there is more than one neuron have the largest input in any cluster, the neuron which has the smallest index will show a non-zero output.

The change of U_{ij} is given by motion equation. It can be shown as:

$$\frac{dU_{ij}}{dt} = -\frac{\partial E}{\partial V_{ij}} \tag{2.15}$$

However, they are approximated in the form of a first order Euler method:

$$\Delta U_{ij} = \frac{dU_{ij}}{dt} \tag{2.16}$$

So the input U_{ij} is based on the first order Euler method, and every neuron of the MNN is iteratively updated by using the following equation:

$$U_{ij}(t+1) = U_{ij}(t) + \Delta U_{ij}(t) \tag{2.17}$$

2.2.4 Elastic Net

The advantage of elastic networks is the geometric nature of the algorithm. It means that the progress of the algorithm can be visually tracked. In addition, the number of neurons needed is scaled linearly with the number of cities. [32]However, although it produces quite good solutions, running times can be large. Many researchers want to improve the elastic network and have presented their methods.

In 1990, Simic gave an idea of using statistical mechanics as the underlying theory which mixed both the Hopfield neural network and the elastic neural network algorithm together. In 1995, Vakhutinsky and Golden introduced a hierarchical strategy to elastic neural network for solving travelling salesman problem. The method of the algorithm is to divide the city area. It made the nodes (cities) into smaller districts then replace the nodes in each district with their "center of gravity".[33][34]

The elastic network was proposed in 1987 as an effective method of travel salesman problem. Later in 1989, Durbin et al. Described, analyzed and evaluated elastomeric webs in detail. Elastic webs can be briefly introduced to ensure their main properties. It has a set of n nodes in the plane within a unit square. And it also has a set of m dynamic points that define a rubber band. The rubber band is shown to be a small circle nearby the center of gravity of the node at the beginning. It then stretches toward the node by tracking the minimum value of the energy function:

$$\mathrm{E} = -\alpha \mathrm{K} \sum_{i=1}^{n} ln \sum_{j=1}^{m} e^{-\|x_i - y_j\|/2k^2} + \beta \sum_{j=1}^{m} \|y_j - y_{j+1}\| \tag{2.18}$$

Where $\mathrm{K} \to 0$ and $\|a-b\| = (a_1-b_1)^2 + (a_2-b_2)^2$. The points' position can define the rubber band. And it is updated through the formula which can be shown as:

$\Delta \mathrm{Y} = -\mathrm{K}\nabla_{\mathrm{Y}} E$ (gradient descent). Where the right side of the equation is set up by two items: the first one is responsible for the point-to-node attraction on the rubber band, and the second is for the shortest trip. We can think of as the temperature, we use some cooling schedule to minimize the energy.

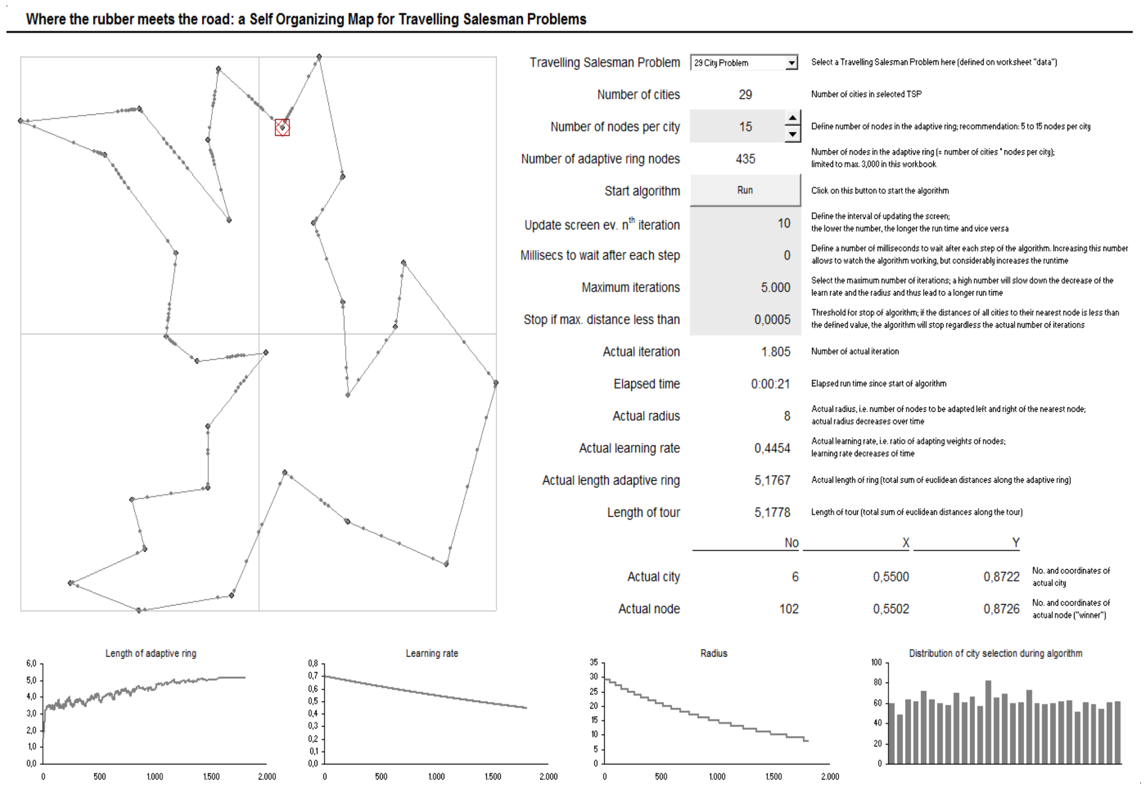


Fig.2.5 Elastic Net for travelling salesman problem

2.2.5 Dendritic Neuron Network

As we all known, the structure of every neuron is unique, it contains three parts: the cell body, dendrite and axon. The dendrite receives the signal from other neurons, then the signal is computed at the synapse and transmitted to the cell body. If the signal into the cell body exceeds the holding threshold, the cell will fire and send the signal down to other neurons through axon.

In 1943, a simple neuron model is proposed by McCulloch and Pitts in which the dendrites and synapses are independents and there are no effects on them from one to another. However, in 1987, Minsky and Papert indicated that the McCulloch-Pitts model is limited to solve complex problems.[35]

Different from the McCulloch-Pitts model which do not consider the dendritic structure in the neuron, Dendritic Neuron Model (DNM model) is proposed in our researches. The DNM model can be generalized as follow:[36]

① The dendrites can be initialized by any arbitrary decision.

② The synapses on the same branch interact each other.

③ The nonlinear interaction produced in a dendrite can express by a logical network.

④ After learning, the branches' ripened number and the locations and types of synapses on the branches will be synthesized.

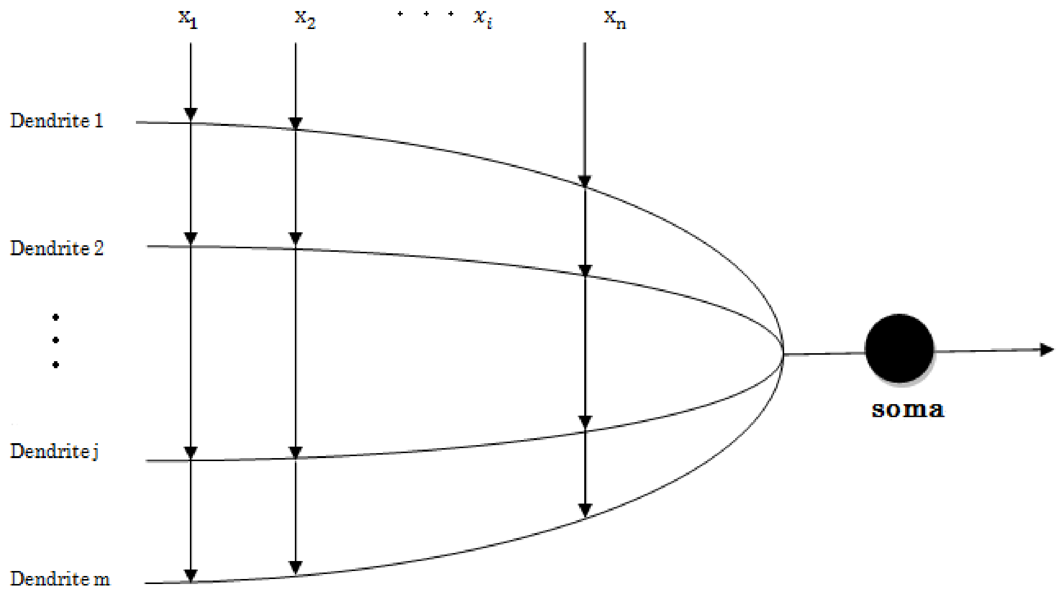


Fig.2.6 Neuron Model with Dendritic Nonlinearity.

As shown in Fig.2.5, the dendritic branches receive signals from x_1 to x_n, then perform a simple multiplication on their own signal. At the junction of the branches, the outputs are summed up and then conducted to soma. If the input of the soma is bigger than a threshold, the cell will fire it, then the cell will send it to other neurons through the axon.

Synaptic Function: In the connection layer, there is a sigmoid function reflects the interaction among the synapses in a dendrite. The output of the synapse whose address is from the $i-\mathrm{th}\,(i=1,2,\ldots,m)$ input to the $j-\mathrm{th}(j=1,2,\ldots,n)$ branch is given by Eq2.19.

$$Y_{ij}=\frac{1}{1+e^{-k(w_{ij}-\theta_{ij})}} \tag{2.19}$$

w_{ij},θ_{ij} respectively means the connection parameters, absolutely, k is a positive constant. When k becomes big enough, the sigmoid function will turn out to be similar to a step function. Through the change of the value of w_{ij} and θ_{ij}, four types of synaptic connections can be defined: a direct connection, an inverted connection, a constant-0 connection(◎), and constant-1 connection(①).

Dendritic Function: It performs a simple multiplication on various synaptic connections of the branch. The output of the j- th branch is given by

$$Z_j = \prod_{i=1}^{n} Y_{ij} \tag{2.20}$$

Membrane Function: It is approximated as follow:

$$\mathrm{V} = \sum_{j=1}^{m} Z_j \tag{2.21}$$

Soma Function: The function of the soma which is introduced by a sigmoid operation, when the k is considered as a positive constant, then the γ can be considered as a threshold from 0 to 1.

$$\mathrm{O} = \frac{1}{1+e^{-k(V-\gamma)}} \tag{2.22}$$

Learning Function: Because DNM is a feed-forward network with continuous functions, the error back-propagation-like algorithm is valid for DNM. By using the learning rule, the error between the target vector and the actual output vector can be expressed as follow:

$$\mathrm{E} = \frac{1}{2}\,(\,T - O\,)^2 \tag{2.23}$$

We should pay attention that the synaptic parameters w_{ih} and θ_{ij} can be modified in the direction to decrease the value of E. And they can be described as:

$$\Delta w_{ij}(t) = -\mu \frac{\partial E}{\partial w_{ij}} \tag{2.24}$$

$$\Delta \theta_{ij}(t) = -\mu \frac{\partial E}{\partial \theta_{ij}} \tag{2.25}$$

Where μ is a positive constant that represents the learning rate. A low learning rate makes the convergence very slow while a high learning rate is difficult for making the error to converge. And the partial differential of E with respect to w_{ih} and θ_{ij} are computed as follow:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial V} \cdot \frac{\partial V}{\partial Z_j} \cdot \frac{\partial Z_j}{\partial Y_{ij}} \cdot \frac{\partial Y_{ij}}{\partial w_{ij}} \tag{2.26}$$

$$\frac{\partial E}{\partial \theta_{ij}} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial V} \cdot \frac{\partial V}{\partial Z_j} \cdot \frac{\partial Z_j}{\partial Y_{ij}} \cdot \frac{\partial Y_{ij}}{\partial \theta_{ij}} \tag{2.27}$$

2.3 Combination model

The linear and nonlinear models on their own problems have gotten great successes. However, they are not suitable for all cases of general model. Bates, J.M., Granger, C.W.J. Said the portfolio model which is with the both modeling ability will be a good alternative when forecasting time-series data. Therefore, this study proposed a combination of model is composed of linear and nonlinear. Therefore, performance can be used to improve overall portfolio model to simulate the linear and nonlinear model. [37][38][39]

As experience, it is reasonable that we can consider time-series is composed of linear autocorrelations and non-linear components which can be performed as:

$$Y_t = L_t + N_t \tag{2.28}$$

Here, the L_t represents the linear component, while the N_t represents the nonlinear component of the combined model. Both L_t and N_t should be estimated for the data set. First, the author let linear model (here we use the Seasonal trend-ARIMA model for the data performs the obvious seasonal trends) to model the linear part, so that the residuals from the linear model will only contain the nonlinear relationship. Make R_t represents the residual at time t, then we can know:

$$R_t = Z_t - \widehat{L_t} \tag{2.29}$$

Where $\widehat{L_t}$ represents the forecast value of the linear model at period t. By modeling residuals using nonlinear model (here we use the DNM model), nonlinear relationships can be discovered. In this paper, we built the model with the following input layers:

$$R_t^{linear} = f^{nonlinear}\left(R_{t-1}^{linear}, R_{t-2}^{linear}, R_{t-3}^{linear}, R_{t-4}^{linear}\right) + e_t \tag{2.30}$$

Where R_t^{linear} is the residual at time t, and it is calculated through the ARIMA model, $f^{nonlinear}$ is a nonlinear function which is decided by the DNM model and e_t is the random error. And the combined forecast can be performed as:

$$\widehat{Y}_t = \widehat{L}_t + \widehat{N}_t \tag{2.31}$$

Where $\widehat{N}_t$ is the forecasting value of Eq 2.30.

3. Time-Series forecasting for House Price Index of China (HPI)

3.1 Introduction of HPI

Interest in the global economy and policy community is growing during the ten-year boom of the property market in China. A major problem is that the collapse of the housing market could undermine the Chinese economy as a result of the housing market bubble, which in turn could have an infectious impact in the United States and Europe. [40]So whether the bubble of Chinese housing market burst or Chinese housing market continues to prosper related to development of not only China, but also the world.

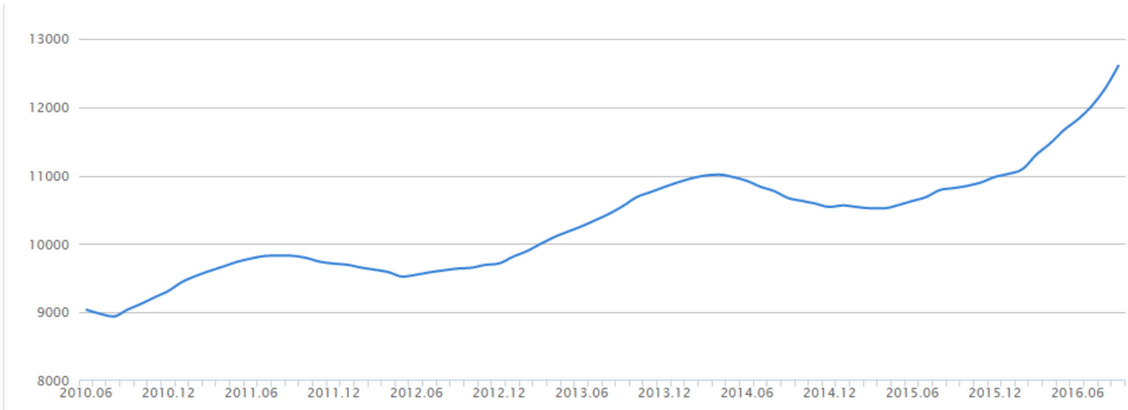


Fig.3.1 HPI of china from 2010 to 2016

In this paper, we will use the house price index of china to forecast the Chinese housing market. House Price Index (HPI) is a set of price indices to reflect the major cities in the country's real estate market conditions and trends in the development of the index system and analysis methods. It uses the weighted average method to calculate the national housing price index as Eq.1 shown:[55]

$$P_j^t = \frac{\sum P_{ij}^t * Q_{ij}}{\sum Q_{ij}} \qquad (1)$$

Where P_j^t represents the average price of the j-th city in period t, P_{ij}^t represents the average price of the j-th city in period t of the i-th project, and

Q_{ij} represents the adjusted construction area of the project. In this paper we used the HPI data which is calculated by China Index Academy. [41,42,43]

In this paper, we used DNM model to fit the HPI data and forecast the trends of Chinese housing market. And then we use the statistical model (the ES model) to test the effectiveness of the DNM model.

3.2 Data set and Prediction

Here we choose the HPI data of China from 1995: 1 to 2012: 9. The collected data were divided into two sets: the training data (data before 2010) and the testing data (data after 2010).[55][56]

As Fig.3.1 shown, the house price index data has extremely trends, then we choose the ES model to fit the data.

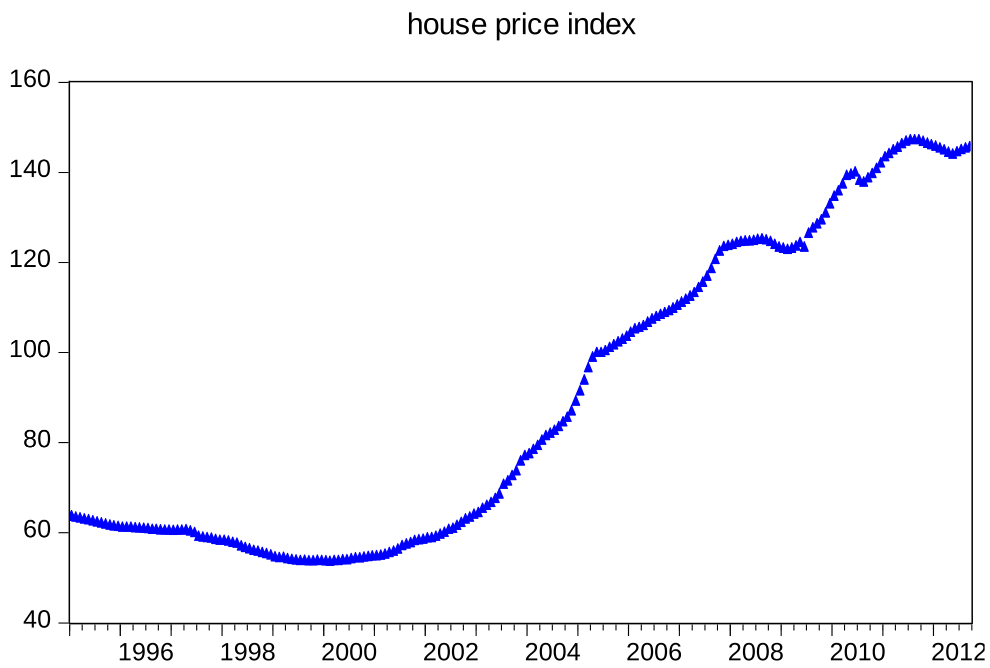


Fig.3.2 the trend of house price index

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.992	0.992	212.48	0.000
		2	0.983	-0.029	422.27	0.000
		3	0.974	-0.025	629.23	0.000
		4	0.965	-0.021	833.25	0.000
		5	0.955	-0.018	1034.2	0.000
		6	0.946	-0.030	1232.0	0.000
		7	0.935	-0.029	1426.5	0.000
		8	0.924	-0.027	1617.4	0.000
		9	0.913	-0.027	1804.6	0.000
		10	0.902	-0.025	1988.1	0.000
		11	0.890	-0.024	2167.7	0.000
		12	0.878	-0.024	2343.3	0.000
		13	0.865	-0.022	2514.8	0.000
		14	0.853	-0.019	2682.1	0.000
		15	0.840	-0.018	2845.1	0.000
		16	0.826	-0.014	3003.9	0.000
		17	0.813	-0.010	3158.4	0.000
		18	0.800	-0.009	3308.6	0.000
		19	0.786	-0.011	3454.5	0.000
		20	0.773	-0.007	3596.1	0.000
		21	0.759	-0.010	3733.5	0.000
		22	0.745	-0.001	3866.7	0.000
		23	0.732	-0.002	3995.8	0.000
		24	0.718	-0.002	4120.8	0.000
		25	0.705	-0.006	4241.8	0.000
		26	0.691	-0.005	4358.8	0.000
		27	0.677	-0.027	4471.8	0.000
		28	0.663	-0.048	4580.6	0.000
		29	0.648	-0.017	4685.2	0.000
		30	0.633	-0.022	4785.6	0.000
		31	0.619	-0.002	4881.9	0.000
		32	0.604	-0.004	4974.1	0.000
		33	0.589	-0.006	5062.5	0.000
		34	0.575	0.004	5146.9	0.000
		35	0.560	0.007	5227.7	0.000
		36	0.546	0.001	5304.9	0.000

Fig.3.3 Test of autocorrelation and partial correlation

In the view of the severe autocorrelation and partial autocorrelation of time-series data, we also checked the HCI data's autocorrelation and partial autocorrelation. As Fig.3.3 shown, there is no significant autocorrelation and partial autocorrelation, and the results are good enough that can lead our research to a ideal state.

As Table 3.1 shown, we summarize the experimental results and choose the best one of DNM model based on the orthogonal array and factor assignment. Here MSD values are calculated by $\bar{x} \pm s$, where $\bar{x}$ means the mean of the results over 20 runs, and s means the standard deviation. It can verify that the data is closer to reality or not. Finally we choose the result of No.10 to compare with the ES model. As Fig. 3.4 show, when we use the DNM model to fit and forecast the data, the error of the data can be quickly converged during the experiment.

Table 3.1 Results based on the orthogonal array and factor assignment of the DNM

No	M	μ	k_{soma}	θ_{soma}	MSD
1	15	0.05	1	0	93.01 ± 33.71
2	15	0.05	3	0.3	90.59 ± 32.54
3	15	0.01	5	0.5	92.41 ± 31.53
4	15	0.01	10	0.9	95.02 ± 33.74
5	25	0.05	1	0	92.08 ± 33.67
6	25	0.05	3	0.3	90.01 ± 32.24
7	25	0.01	5	0.5	91.35 ± 32.23
8	25	0.01	10	0	93.12 ± 32.24
9	30	0.05	1	0.9	90.73 ± 31.96
10	30	0.05	3	0.3	89.39 ± 31.34
11	30	0.01	5	0.5	90.55 ± 33.53
12	30	0.01	10	0.9	91.09 ± 32.94

* M means number of dendrites

As Fig.3.5 – Fig.3.7 shown, the DNM model plays as well as the popular traditional statistical model (ES model) in forecasting the time-series data. Further attesting to this, some quantitative statistical metrics are used to evaluate the performance of models. They are NMSE (normalized mean square error), APE (absolute percentage of error), and R (correlation coefficient). When the values of NMSE and APE is small enough, we can regard the predicted values is as close as the actual values. The metric R is used to calculate the correlation of the actual and the predicted values. The calculation of them is set up in Table 3.2. The result is shown as Table 3.3.

Table 3.2 Calculation of the NMSE, APE and R

Metrics	Calculation
NMSE	NMSE$=\frac{\sum_{i=1}^{n}(a_i-b_i)^2}{n\delta^2}$; $\delta=\frac{\sum_{i=1}^{n}(a_i-\bar{a})^2}{n-1}$
APE	APE$=\frac{\sum_{i=1}^{n}\lvert(a_i-b_i)/a_i\rvert}{n}\times 100\%$
R	R$=\frac{\sum_{i=1}^{n}(a_i b_i)}{\sqrt{\sum_{i=1}^{n}a_i^2}\times\sqrt{\sum_{i=1}^{n}b_i^2}}$

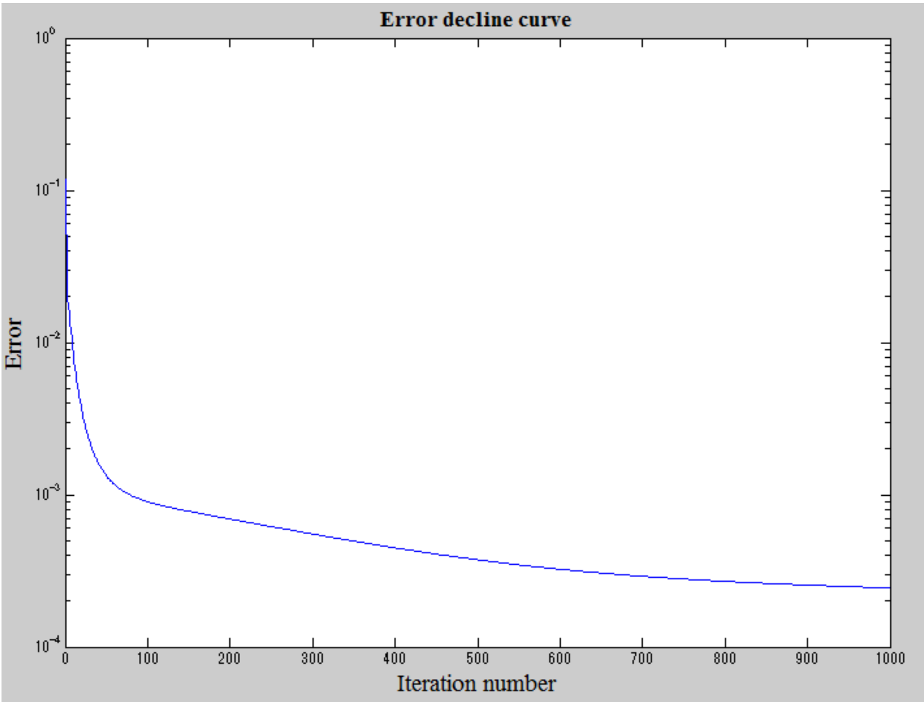


Fig.3.4 Error decline curve of the DNM model

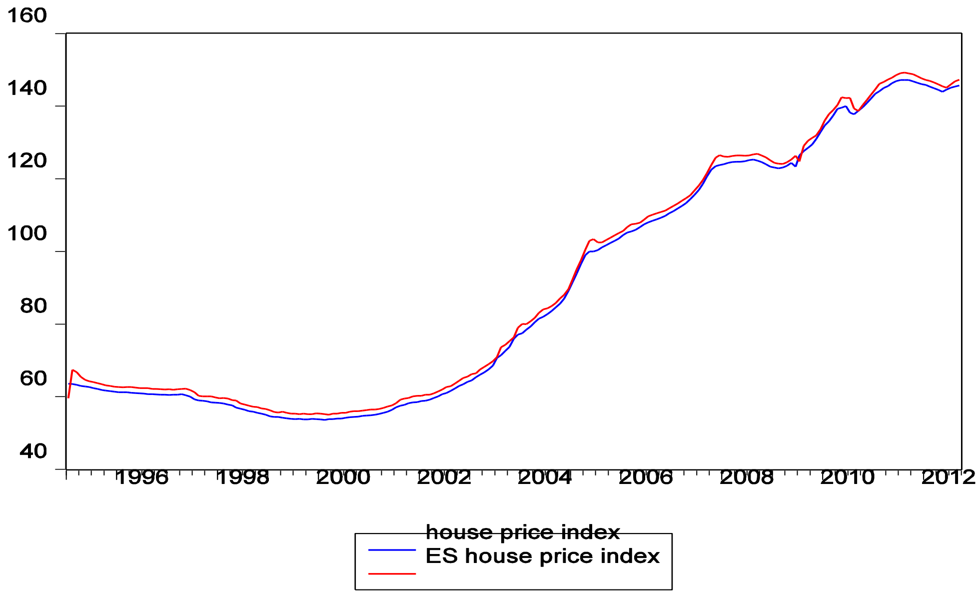


Fig.3.5 Data fit of the ES model

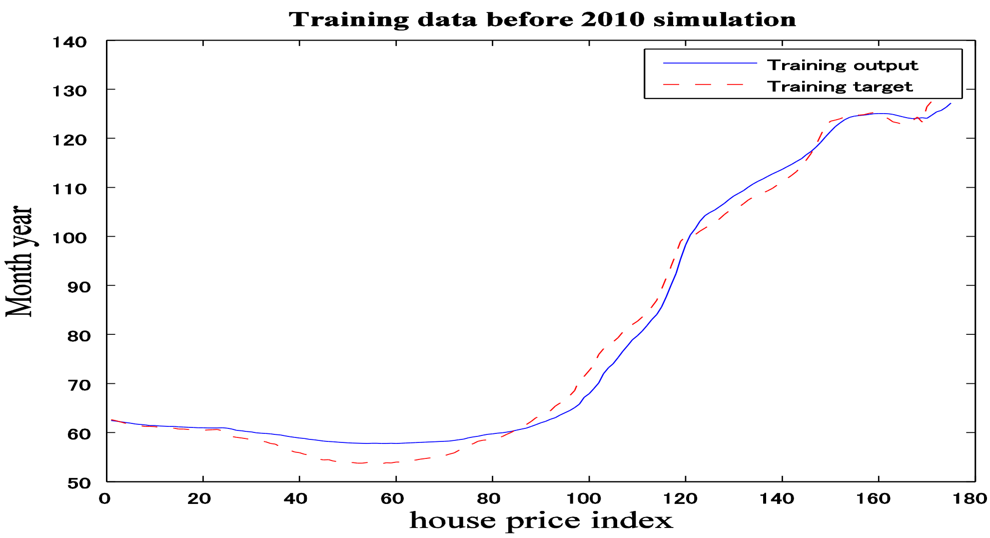


Fig.3.6 Data fit of the DNM model

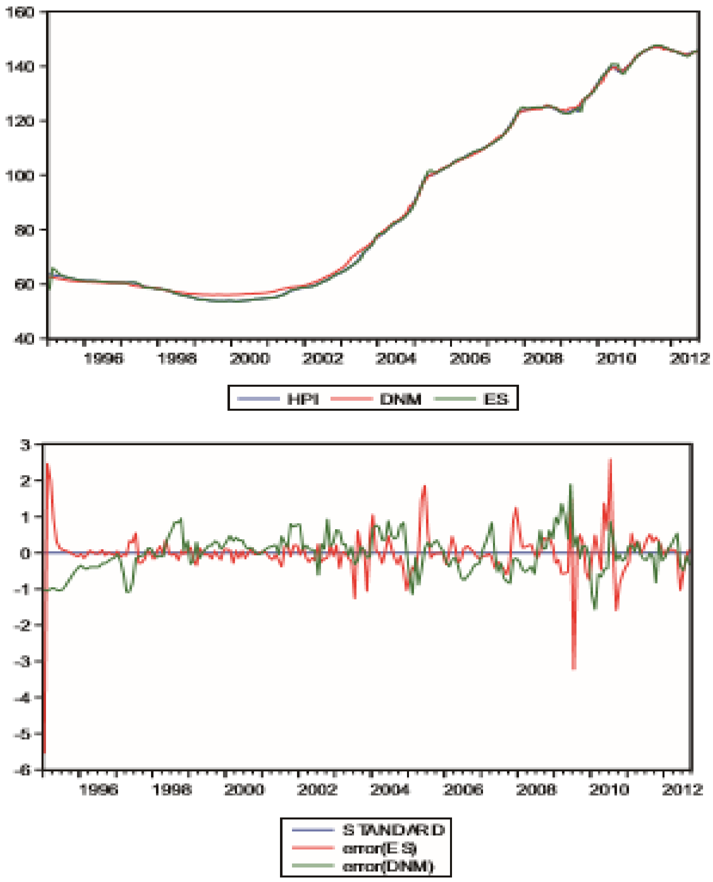


Fig.3.7 Data fit and residual error of the DNM model and the ES model

Table 3.3 The compared results of the DNM model and the ES model

Metrics	The DNN model	The ES model
NMSE	0.279	0.322
APE	0.79	0.83
R	0.92	0.90

3.3 Conclusions of experiment (HPI)

In this study of forecasting HPI of China, we proposed our model, the DNM model can forecast the HCI data very effectively. We use the HCI data which we collected from China Index Academy to fit the DNM model. The results showed the DNM model performed very well in fitting and forecasting the HCI data. Then we verified the effectiveness of our model by comparing the ES model and got the expected result.

The contributions of our study is that it is based on neuron model with dendritic nonlinearity model and it theoretically strengthens the assumption that a neural network model performs effectively not only in the nonlinear model but also in the time-series data.

In our further work, to excavate more efficient of the DNM model, we are considering to mix the DNM model and the statistical model together to create a combination model which maybe predict random data much better.

4. Time-Series Forecasting for Tourism Economy

4.1 Introduction of Tourism Economy

With the impact of Global Internationalization, Tourism is also in a state of rapid development. As we all know, Tourism to a country's economic and social development is huge. It not only can be used for commercial, trade and capital investment, and can also create employment and entrepreneurship for labor, protection of heritage and cultural value (as Table 4.1 and Fig.4.1 shown). Every country wants to know the data of inbound tourists, in order to select the appropriate strategy for its economic welfare. Therefore, the reliable prediction is needed, and plays an important role in tourism planning. [44][45]

Table 4.1 Inbound tourism consumption

(Billion Yen)

Products	same-day visitors	touris ts	total visitors
Characteristic products	0	1167	1167
Accommodation services	0	496	496
Food and beverage servicing services	0	303	303
Passenger transport services	0	328	328
Travel agency, tour operator and tourist guide services	0	8	8
Cultural services	0	10	10
Recreation and other entertainment services	0	8	8
Miscellaneous toirism service	0	14	14
Connected products	0	483	483
TOTAL	**0**	**1650**	**1650**

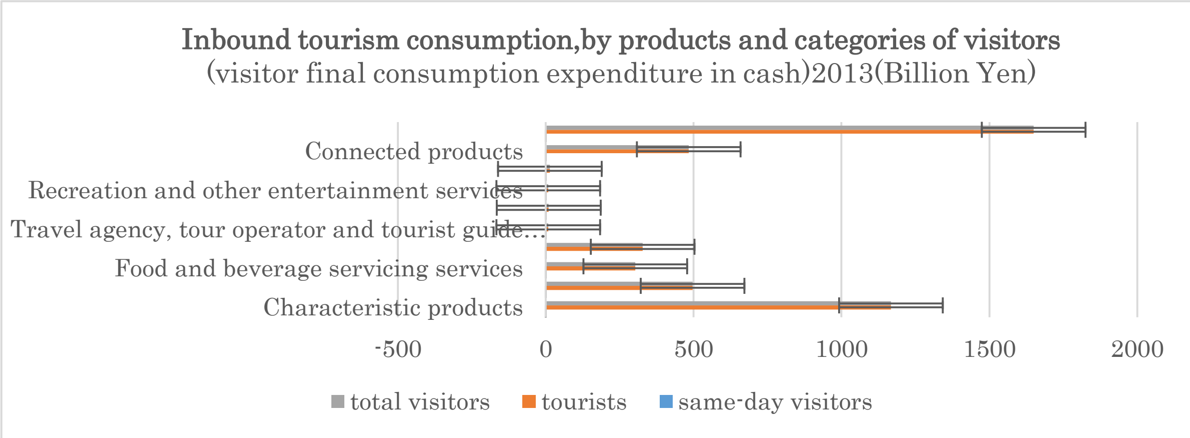


Fig.4.1 Inbound tourism consumption

Accurate forecasting lays the foundation for better tourism planning and management. Therefore, more effective tourism demand research and forecasting technology is needed.[46]

Over the past two decades, tourism demand modeling and forecasting that is one of the most important areas in tourism research has called for more and more attention of both Theoretical scholars and practitioners. As Song, and Li concluded, Twenty years ago, only a few academic journals which have published travel-related research. However, after two decades, there are now over 70 journals which are serving a thriving research community, covering more than 3,000 universities on five continents. Unfortunately, there is no panacea for tourism demand prediction. [47][48]

In recent years, statistical data has been widely used in the study of tourism economy. In statistical methods, time-series prediction is an important field of prediction. It can be divided into two categories: linear method and nonlinear method. The most popular linear methods are the naive models, the exponential smoothing (ES) model and the autoregressive integral moving average model (ARIMA). And among them, the most advanced linear method prediction model is the ARIMA model, and it has gotten great successes in many practical applications. If the linear model is a good approximation to the underlying data generation process, they can be considered as the preferred model. However, more complex nonlinear models should be considered if linear models are not well implemented in both intra-sample and out-of-sample predictions. While there is some doubt about the prediction of travel demand based on neural networks, nonlinear methods are generally considered superior to linear methods. For they have gotten great success in modeling economic behavior which can help to make better decisions. [49]

After decades of development, it has formed hundreds of artificial neural networks. In 1974, P.Werbos firstly proposed learning algorithm for multi-layer network in his doctoral thesis, but the algorithm has not been enough attention and wide range of applications. Until the 20th century, mid-1980s, David Rumelhart, Geoffrey Hintorl and RorlaldWilliams, David Parkr, and Yannn Le Cun each independently discovered the BP algorithm. In 1986, California's PDP (parallel distributed processing) group issued a book called "Parallel Distributed Processing", from then on, the BP algorithm was applied as a study in the neural network. BP neural network is the network which is trained after this algorithm.[50]

Up to now, there are many researchers using a lot of methods to forecast the tourism demand. And they can be divided into three types: time series, neural network and combined models. In 2014, J.P. Terxeira and P.O. Fernandes published [25 years of time series forecasting, International Journal of Forecasting], in which, the three methods are all be mentioned. Except those, there are also a lot of authors using the three methods separately. Such as Box & Jenkins, Cho, Chu, Song & Li, Law, Qu & Zhang, Shahrabi &Hadavandi & Asadi, Li & Pan & Law & Huang, Kawakubo & kubokawa have used the traditional time series methods to forecast the tourism demand. As neural network is widely known, there are many authors turning to use the neural network to forecast the time series data such as Chen & Lai & yeh, Claveria & Torra, Davies & Petruccelli &Pemberton, Constantino & Fernanded & Teixeira, Law, Lin & Chen & Lee, Pai &Hong. With the progress of science, more and more methods are being used. The combined models are the most popular method in them. And up to now, Bates & Granger, Chen, K., Chen Kuan-Yu, Shen & Li & Song, Yan have used this method and got the expected results. Besides these, Lin & Pai & Lu & Chang, Pai & Hung &Lin also proposed the different methods like support vector regression and novel hybrid system.[51]

In this sample, we mix the most advanced linear model (SARIMA model) with the innovative neural network model (DNM model) together and call the mixed model the SA-D model, By comparing, we obtained the SA-D model is much better than the DNM model to tourism demand forecasting whether in the time used or the accuracy of prediction.

4.2 Data set and Prediction

As a result of the rapid economic growth and international tourism promotion, visitors to Japan increased year by year. Here we choose the inbound tourists from 2009:1 to 2015:12. And the process of data set is shown in Table.2. The collected data were divided into two sets: the training data (data before 2015) and the testing data (data of 2015). [53][54]

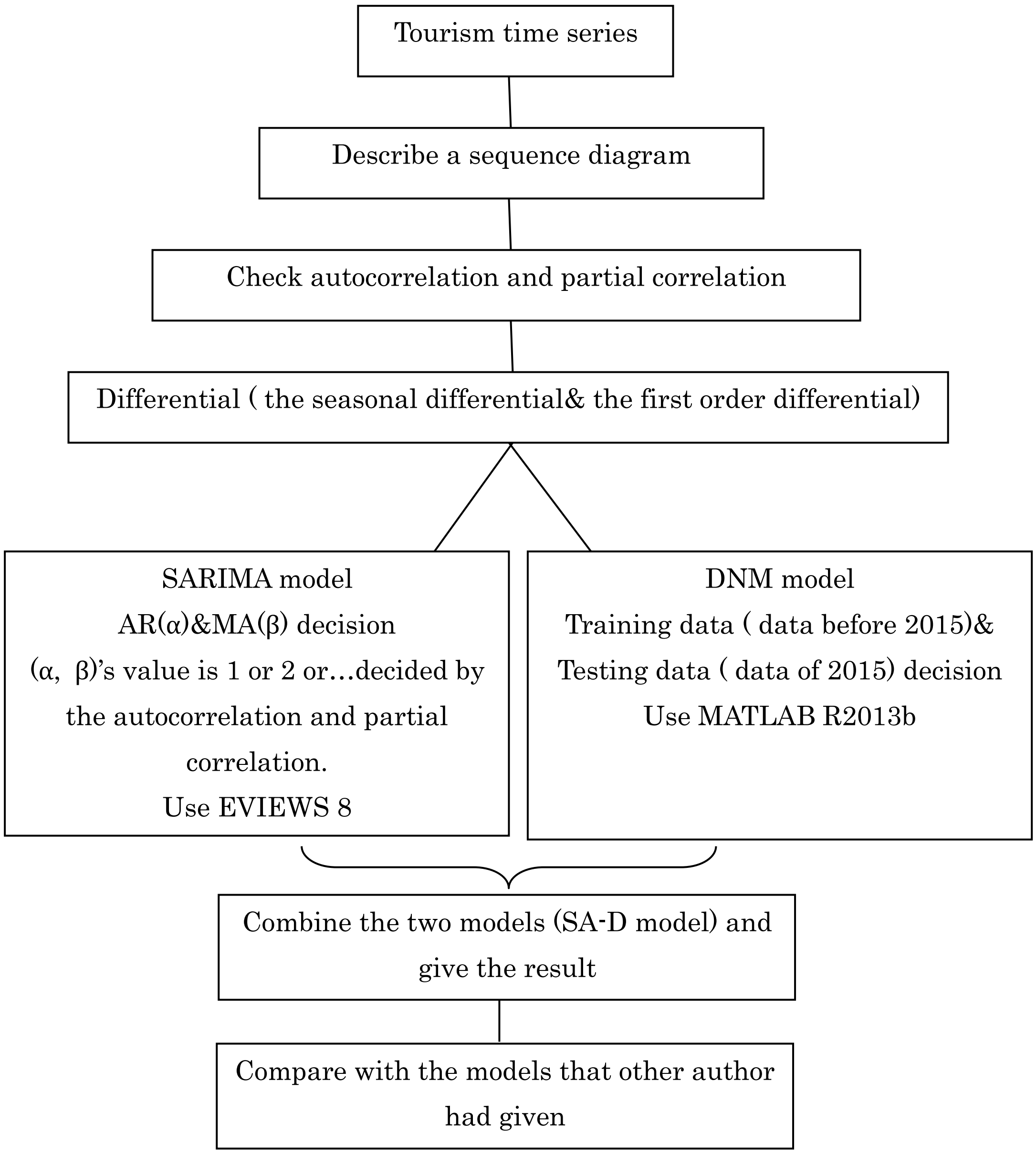


Fig.4.2 Process of data set.

Performance criteria

The predictive performance of the predictive model is evaluated using a number of quantitative statistical measures such as normalized mean square error (NMSE), absolute error percentage (APE), R (correlation coefficient), and program run time (PRT) (Table 4.2). NMSE and APE are used to measure the deviation between the predicted and actual values. The smaller the value of NMSE and APE, the closer the predicted value is to the actual value. The measure R is used to measure the correlation between the actual value and the predicted value. PRT can measure the running speed of the model.

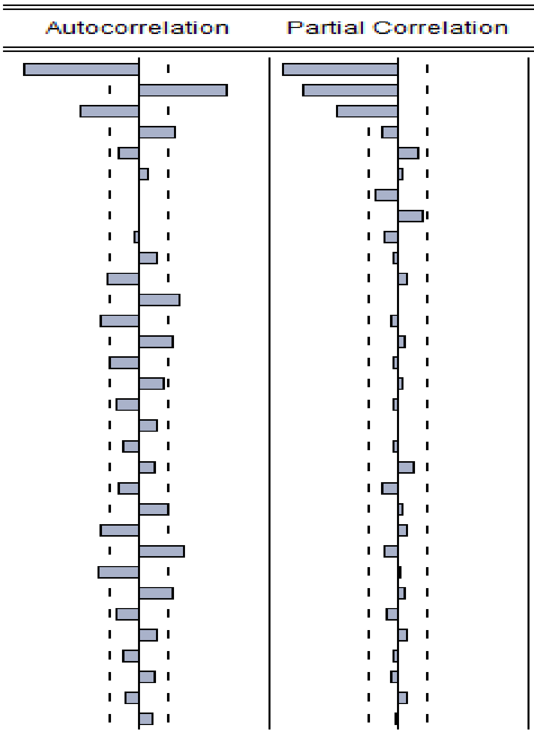
Table 4.2 Calculations of the performance metrics.

Metrics	Calculation
NMSE	NMSE=$\frac{\sum_{i=1}^{n}(a_i-b_i)^2}{n\delta^2}$; $\delta = \frac{\sum_{i=1}^{n}(a_i-\bar{a})^2}{n-1}$
APE	APE=$\frac{\sum_{i=1}^{n}\lvert(a_i-b_i)/a_i\rvert}{n} \times 100\%$
R	R=$\frac{\sum_{i=1}^{n}(a_i b_i)}{\sqrt{\sum_{i=1}^{n} a_i^2} \times \sqrt{\sum_{i=1}^{n} b_i^2}}$
PRT	Decided by the actual operation.

* a_i and b_i are the actual values and the predicted values

Experimental results

For the data has significant seasonal changes periodically. We use the SARIMA model in this paper to eliminate the linear trend. As the Fig.4.3 shown, we can decide the possible generations of the ARIMA model, and use the Akaike Information Criterion (AIC) to test which of the generations is the best.



	AC	PAC	Q-Stat	Prob
1	-0.911	-0.911	62.260	0.000
2	0.701	-0.756	99.652	0.000
3	-0.468	-0.492	116.59	0.000
4	0.283	-0.131	122.87	0.000
5	-0.156	0.169	124.80	0.000
6	0.067	0.040	125.16	0.000
7	-0.007	-0.188	125.17	0.000
8	-0.007	0.203	125.17	0.000
9	-0.045	-0.098	125.34	0.000
10	0.150	-0.041	127.29	0.000
11	-0.259	0.074	133.14	0.000
12	0.320	-0.003	142.22	0.000
13	-0.317	-0.057	151.30	0.000
14	0.276	0.044	158.29	0.000
15	-0.231	-0.030	163.28	0.000
16	0.201	0.038	167.12	0.000
17	-0.179	-0.035	170.21	0.000
18	0.153	0.008	172.51	0.000
19	-0.127	-0.040	174.12	0.000
20	0.122	0.116	175.64	0.000
21	-0.160	-0.137	178.32	0.000
22	0.238	0.035	184.35	0.000
23	-0.319	0.081	195.39	0.000
24	0.357	-0.102	209.54	0.000
25	-0.332	0.022	222.05	0.000
26	0.262	0.050	230.02	0.000
27	-0.189	-0.088	234.23	0.000
28	0.142	0.079	236.68	0.000
29	-0.126	-0.032	238.63	0.000
30	0.120	-0.051	240.47	0.000
31	-0.110	0.070	242.03	0.000
32	0.099	-0.017	243.33	0.000

Fig.4.3 Autocorrelation and Partial correlation.

Through the SARIMA model, we get the data that has no linear trend, and train the data separately by the DNN model and the SA-D model. We can get the results of the DNM model and the SA-D model as follow.

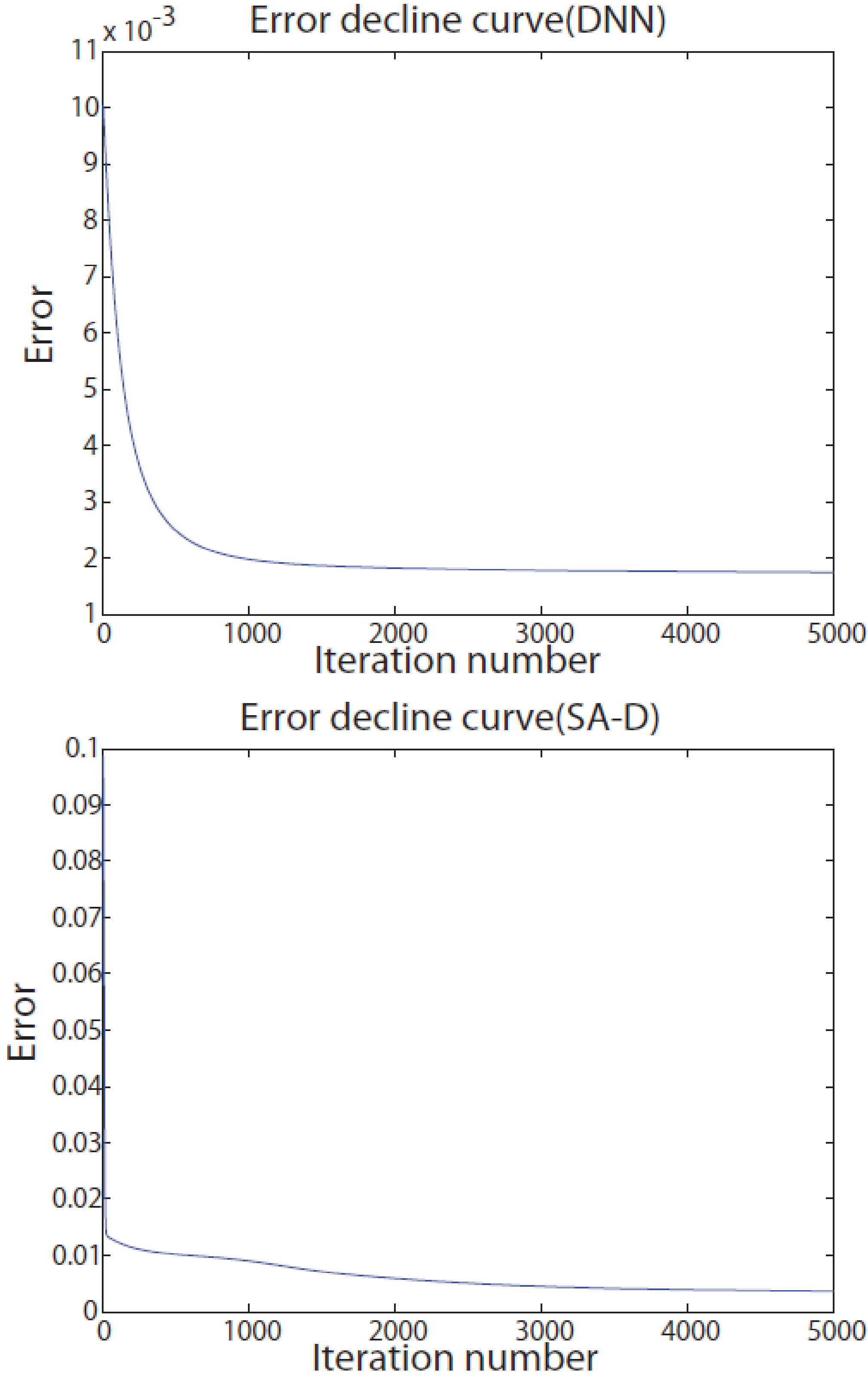
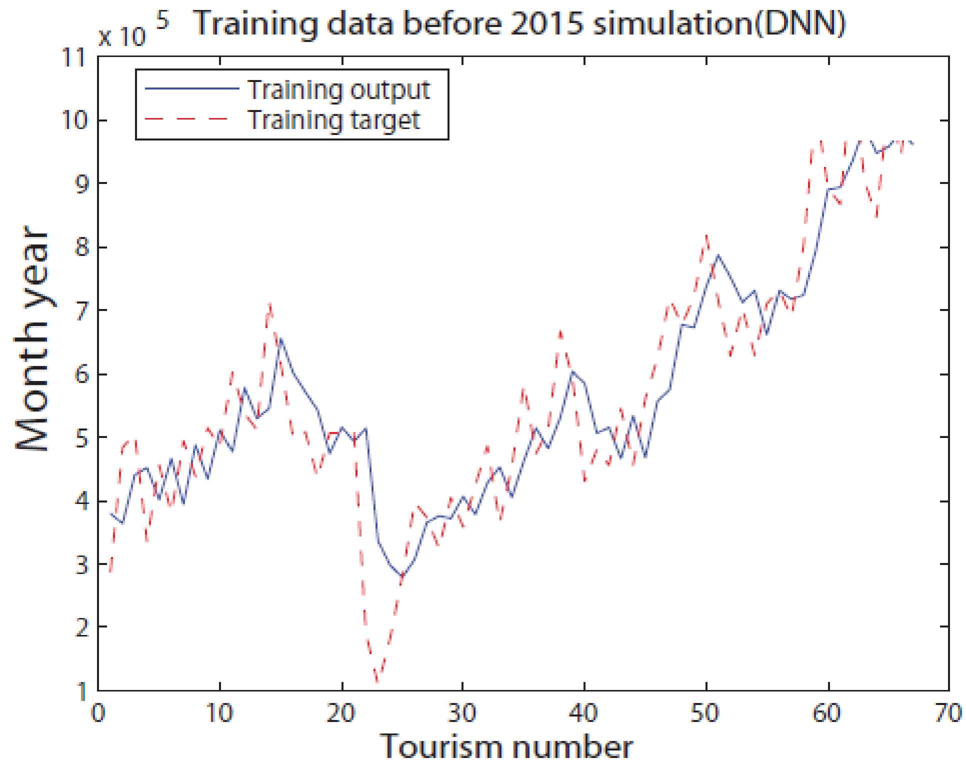


Fig.4.4 Error decline curve of the DNM model and the SA-D model



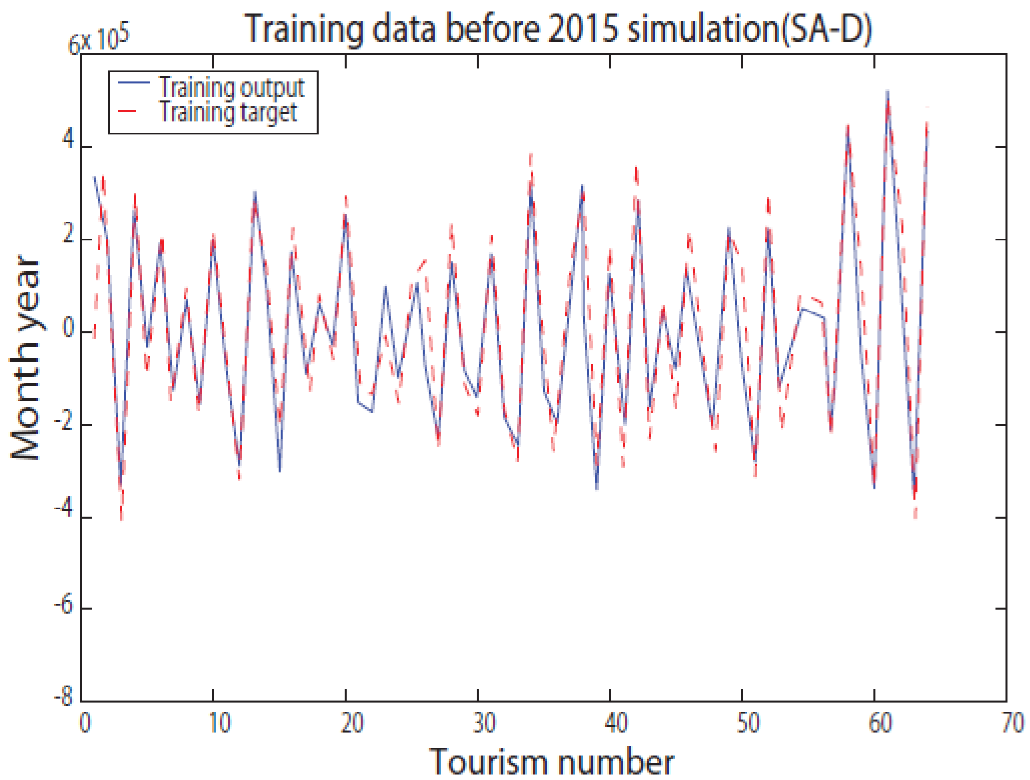


Fig.4.5 Training data before 2015 simulation of the DNM model and the SA-D model

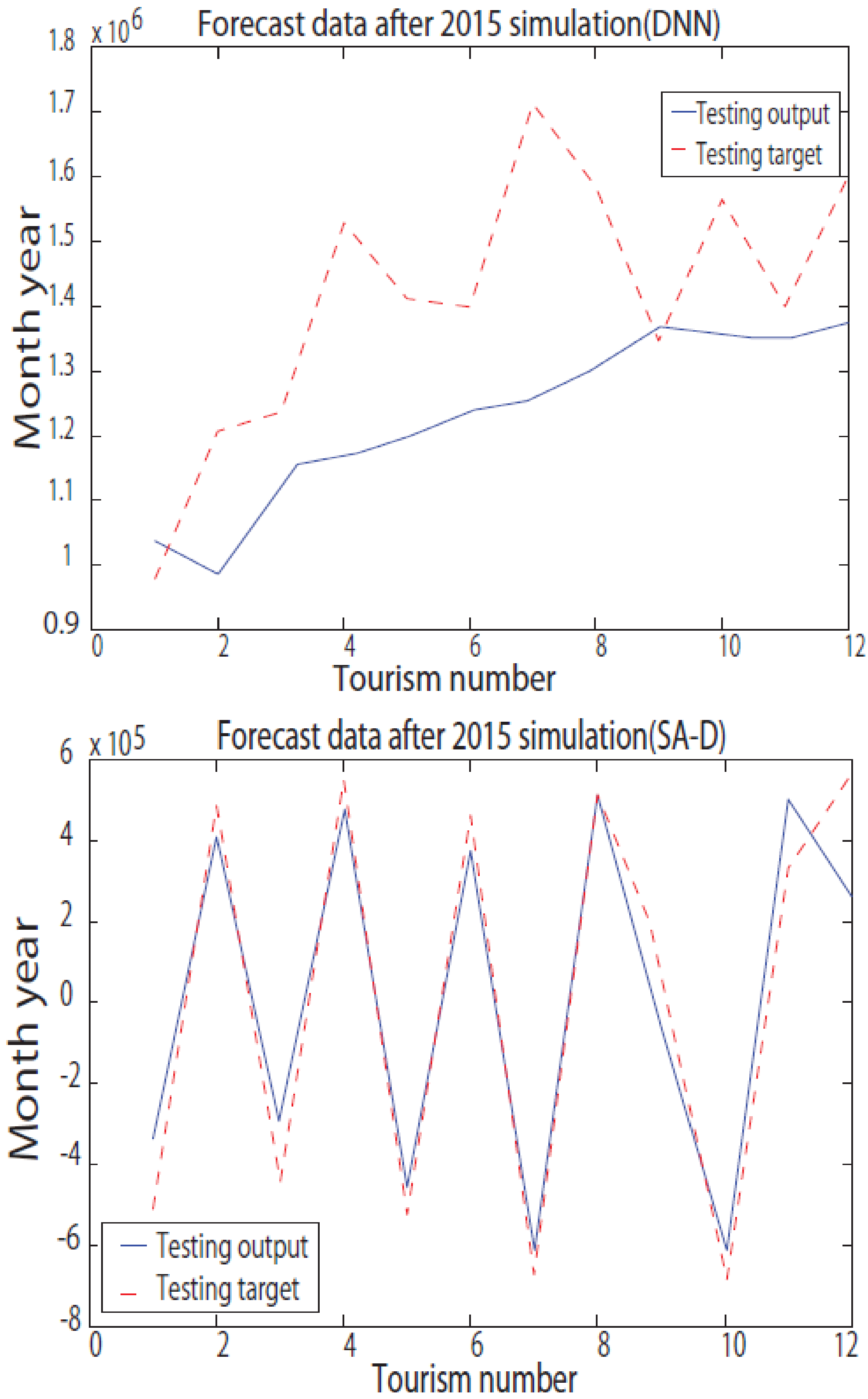


Fig.4.6 Forecast data after 2015 simulation of the DNM model and the SA-D model

As Fig.4.4 to Fig.4.6 shown, we can see that the results of the SA-D model perform much better than those of the DNM model. In order to deeply evaluate the performance of the DNM model and the SA-D model, we calculate MAPE, NMSE and R of the testing data set as Table 4.3 shown.

Table 4.3 the compared results of the DNM model and the SA-D model.

Metrics	The DNM model	The SA-D model
NMSE	2.245	0.219
APE	0.87	0.78
R	0.32	0.89
PRT	The DNN model is rapider than the SA-D model	

We can see that although the PRT of the DNM model is rapider than that of the SA-D model, the NMSE, APE and R of the SA-D model are much better than those of the DNM model.

4.3 Models comparing

To demonstrate the validity of the SA-D model, we train the same data that other author had used in the other combination models. And compare the results of the SA-D model and the models other author had proposed. [10] We collected the monthly outbound tourism data that from Taiwan travel to 3 areas (Americas, Europe and Oceania) from Tourism Burean M.O.T.C. Republic of China (Taiwan) [72]. The study time ranges from January of 1998 to June of 2009. The collected data were divided into two parts, training data (data from 1998 to 2007) and testing data (data after 2007) for each tourism demand time series. The author scaled the data within the range of (0, 1) through the following formula.[52]

$$\frac{X_t - X_{min}}{X_{max} - X_{min}} \times 0.7 + 0.15 \qquad (3.1)$$

So we use the data with the same pre-set as the author did and without the data pre-set separately and get our experimental results.

Before comparing with the models, we summarize the experimental results and choose the best one of the SA-D model based on the orthogonal array, factor assignment and statistical tests as Table 4.4 shown. Here the MSD values are calculated by $\bar{x} \pm s$, where $\bar{x}$ means the mean of the results over 20 runs, and s means the standard deviation. It can verify the data is closer to reality or not. And the p value can determine whether the residual is white noise sequence or not after the statistical test by using QLB statistic. Finally we choose the result of the No.7 to do the comparison.

Table 4.4: Results based on the orthogonal array factor assignment and statistical tests of the SA-D model

No	M	μ	k_{soma}	θ_{soma}	MSD	p
1	15	0.05	1	0	0.401 ± 0.169	0.1938
2	15	0.05	3	0.3	0.386 ± 0.170	0.2013
3	15	0.01	5	0.5	0.391 ± 0.171	0.1854
4	15	0.01	10	0.9	0.389 ± 0.167	0.191
5	25	0.05	1	0	0.392 ± 0.165	0.2563
6	25	0.05	3	0.3	0.395 ± 0.164	0.2742
7	25	0.01	5	0.5	0.398 ± 0.161	0.3011
8	25	0.01	10	0	0.390 ± 0.168	0.2916
9	30	0.05	1	0.9	0.402 ± 0.172	0.1928
10	30	0.05	3	0.3	0.399 ± 0.171	0.1897
11	30	0.01	5	0.5	0.394 ± 0.168	0.2001
12	30	0.01	10	0.9	0.397 ± 0.170	0.1936

* M means number of dendrites;

Table 4.5 Comparison of the SA-D model and the other combination models

		Americas	Europe	Oceania
ARIMA+BPNN	APE	13.41	12.95	13.46
	NMSE	0.3992	0.8153	0.5327
	R	0.9918	0.9917	0.9856
ARIMA+SVR	APE	11.46	11.37	11.87
	NMSE	0.2878	0.6316	0.5102
	R	0.9923	0.9917	0.9871
The SA-D model (with data pre-set as other author did)	APE	9.61	9.73	9.89
	NMSE	0.2788	0.4561	0.4968
	R	0.9934	0.9921	0.9864
The SA-D model (without data pre-set)	APE	10.34	10.51	10.87
	NMSE	0.3458	0.5619	0.6027
	R	0.9912	0.9906	0.9891

As table 4.5 shown, our model performed much better results than other author's models. But there have to say that the data pre-set by formula 14 made the results better and reduced the running time of program.

4.4 Conclusions of experiment (Tourism Economy)

We proposed a new model, the SA-D model which mixed the SARIMA model and the DNN model together. First, we used the data that collected from Japan Tourism Agency Ministry of Land, Infrastructure, Transport and Tourism and Japan National Tourism Organization to compare the SA-D model and DNN model, the results showed the SA-D model performed much better in fitting and forecasting the time series data. Then we verified the effectiveness of our model by comparing other author's models and got the expected result.

The contributions of this study lie in two aspects. Our study is based on neuron model with dendritic nonlinearity model and it theoretically strengthens the assumption that neural network model implementation which performs effectively not only in the nonlinear model but also in the time-series data.

This study which mixed the linear model and the nonlinear model together opens the door to further combination models with different methods and models.

5. Conclusion

5.1 Conclusions of the thesis

In my paper, I have presented the concept and the particularity of the time-series firstly. When analyze a time-series data, we should follow the steps as:

① What lets time-series data particular?
② How to loading and processing the time-series data.
③ How to examine smoothness of a time-series data?
④ How to turn a time-series data to be smooth?
⑤ Prediction of time-series data.

After statistics, neural network, and the relationship between them are stated, some statistical models (such as the naïve model, the ES model, and the ARIMA model), neural network models (such as Hopfield neural network, self-organizing map, maximum neural network, and elastic net), and combinatorial model are proposed. The main goal is to study neuron model with dendritic nonlinearity model, and based on it, effective combinatorial model is to be created to fit and forecast the time-series data.

In chapter 2, the statistical models, neural network models and combinatorial model are detailed descripted. The naïve model is the most cost-effective predictive model and provides a benchmark against which more complex models can be compared. This prediction method only applies to time series data. ES model is the most common time series forecasting method. It was developed on the basis of mobile averaging technology. It predicts the effect of the closest actual value on the predicted value which is based on the current actual value and the currently predicted next value, but does not require the quality of past values. The ARIMA is the most welcomed linear model for predicting time-series data. It has been a great success not only in academic research but also in industrial and economic applications. Box and Jenkins suggested that in 1976 the ARIMA model should use at least 50 or preferably 100 observations. Due to the theoretical limitations of the network structure, the importance of different Hopfield networks in practical applications is limited. But in some cases, they may form interesting models. Hopfield networks are often used to classify vectors with binary patterns. The self-organizing map is an artificial neural network. It is trained by using unsupervised learning which generates a low-dimensional, and discretes representation of the input space of the training sample, it is named as a mapping. This mapping is designed to preserve the topological properties of the input space. There are two NP-complete optimization problems, they are called module orientation problems and objective functions, such as bus length and subgraph size, the effectiveness of the largest neuron model, and no constraints imposed. The advantage of elastic networks is that the geometric nature of the algorithm means that the progress of the algorithm can be visually tracked. In addition, the number of neurons needed is scaled linearly with the number of cities. However, although it produces quite good solutions, running times can be large. Dendritic neuron model can be generalized as: the dendrites can be initialized by any arbitrary decision, the synapses on the same branch interact each other, the nonlinear interaction produced in a dendrite can express by a logical network, after learning, the branches' ripened number and the locations and types of synapses on the branches will be synthesized. Both linear and nonlinear models have maintained great successes in theoretical and practical problems. However, none of them is an universal model that can hand all situations. So the combinatorial model proposed in this paper is composed of the linear and nonlinear component can model linear and nonlinear patterns with improved overall forecasting performance.

In chapter 3 and 4, two experimental results are stated. One is the time-series forecasting for house price index of China, in which, the DNM model is proposed, and the ES model is used to be compared with the DNM model which showed its feasibility. The other one is the time-series forecasting for tourism economy, in which a new model is proposed, it is called the SA-D model which mixed the SARIMA model and the DNM model together. The results showed the SA-D model performed much better in fitting and forecasting the time-series data. Then I verified the effectiveness of the SA-D model by comparing other author's models and got the expected result.

5.2 Contributions of the thesis

The contributions of my paper can be divided into three parts:

- I systematically enumerated and stated the time-series forecasting models, and analyzed their advantages and disadvantages respectively.
- My research is based on neuron model with dendritic nonlinearity model and it theoretically strengthens the assumption that neural network model implementation which performs effectively not only in the nonlinear model but also in the time-series data.
- To excavate more efficient of the DNM model, I am considering to mix the DNM model and the statistical model together to create a combinatorial model which maybe predict random data much better, and opened a door to the combinatorial models.

5.3 Suggestion for future research

In recent years, time-series have been receiving considerable attention and prediction approaches are slowly being developed. However, there are still some areas of future research should be developed as soon as possible. Concern with the time-series, approaches of the future research in my opinion are as follows:

Research and development of multivariate methods focuses on a more practical proposition can be made. Some researchers have noted that much work has not been done on multiple time-series models, including multivariate exponential smoothing. I suspect that there are two reasons: one is the lack of multi-variable model of the robust prediction algorithm of empirical research; the other is the lack of easy to use software. Some of the proposed methods are very difficult to estimate for a large number of parameters involved. When others, for example, the multiple exponential smoothing, do not get enough theoretical attention to prepare for everyday use. So I doubt it will be more widely used in the near future.

The prediction method based on non-linear model needs more in-depth research. The development model selection process makes good use of data and prior knowledge, needs to specify forecasting goals, and develops predictive systems that meet these goals.

These areas are still calling for the notices and I believe that in the future researches, they will create techniques to solve the problems of the time-series forecasting.

References

[1] Weizhong Yan, Toward Automatic Time-Series Forecasting Using Neural Networks, IEEE Transactions on neural networks and learning systems 7(2012) VOL23.
[2] Teixeira, J., Fernandes, P., Tourism time series forecast different ANN architectures with time index input, Procedia Technology 5(2012)445-454.
[3] Box, G.E.P., Jenkins, G.M., Time series analysis: Forecasting and Control, San Francisco: Holden-Day (1976).
[4] Burger, C.J.S.C., Dohnal, M., Kathrada, M., Law, R., A practitioners guide to time-series methods for tourism demand forecasting --- A case study of Durban, South Africa, Tourism Managemant 22(2001) 403-409.
[5] B.Widrow, Adaptive Sampled-Data Systems – A Statistical Theory of Adaptation, IRE WESCON Convention Record 4(1959) 74-85.
[6] Davies, N., Petruccelli, J.D., Pemberton, J., An automatic procedure for identification, estimation and forecasting self exciting threshold autoregressive models, Statistician 37(1988) 199-204.
[7] Jan G. De Gooijer, Rob J. Hyndman, 25 years of time series forecasting, International Journal of Forecasting 22(2006) 443-473.
[8] Anderson, J.A., A simple neural networks generating interactive memory, Math. Biosc 14(1972) 197-220
[9] B.Widrow, M.E.Hoff, Associative Storage and Retrieval of Digital Information in Networks of Adaptive 'Neurons', Biological Prototypes and Synthetic Systems 1(1962) p160.
[10] David E. Rumelhart, James L. McClelland, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, MIT Press (1986).
[11] David E. Rumelhart, James L. McClelland ,PDP Research Group, Parallel Distributed Processing, Computer Science and Intelligent System(1986).
[12] Hebb, Donald O., The Organization of Behavior: A Neuropsychological Theory, Wiley, New York, 1949.
[13] McCulloch, W.S., Pitts, W., A logical calculus of the ideas immanent in nervous activity, Mathematical Biophysics 5(1943) 18-27.
[14] W.S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, The bulletin of mathematical biophysics 5(1943) 115-133.

[15] Haiyan Song, Gang Li, Tourism demand modelling and forecasting—A review of recent research, Tourism Management 29(2008) 203-220.
[16] J.J.Hopfield, D.W.Tank, Neural computation of decisions in optimization problems, Biological Cybernetics 52(1985) 141-152.
[17] J.J.Hopfield, D.W.Tank, Computing with neural circuits: a model, Science 233(1986) 625-633.
[18] K.C.Lee, N.Funabiki, Y.Takefuji, A parallel improvement algorithm for the bipartite subgraph problem, IEEE Trans. Neural networks 3(1992) 139-145.
[19] Law, R., Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting, Tourism Management 21(2000) 331-340.
[20] Kohonen, T., Correlation matrix memories, IEEE Trans. Comp., Vol.C-21(1970) 353-359.
[21] M.L. Minsky, S.A., Papert, Perceptrons-Expanded Edition: An introduction to Computational Geometry, MIT press Boston, MA (1987).
[22] Pai, P., Hong, W., An improved neural network model in forecasting arrivals, Annals of Tourism Research 32(2005) 3691-3702.
[23] P.D.Simic, Statistical Mechanics as the underlying theory of 'Elastic' and 'Neural' Optimisations, Network: Computation in neural Systems 1(1990) 88-103.
[24] G.V.Wilson, G.S.Pawley, On the stability of the travelling salesman problem algorithm of Hopfield and Tank, Biological Cybernetics 58(1988) 63-70.
[25] H.A. Constantino, P.O. Fernandes, J.P.Teixeira, Tourism demand modelling and forecasting with artificial neural network models: The Mozambique case study, TEKHNE – Review of Applied Management Studies 12 (2016).
[26] Erwin, E., Obermayer, K., Schulten, K., Self-organizing maps: Stationary states, metastability and convergence rate. Biological Cybernetics 67(1992) 35-45.
[27] Fong-Lin Chu, Using a logistic growth regression model to forecast the demand for tourism in Las Vegas, Tourism Management perspectives 12(2014) 62-67.

[28] J.P. Teixeira, P.O. Fernandes, Tourism time series forecast with artificial neural networks, TEKHNE – Review of Applied Management Studies12 (2014) 26-36.
[29] N. Aras, B.J. Oommen, I.K.Altinel, Kohonen network incorporating explicit atatistics and its application to the travelling salesman problem, Neural Networks 12(1999) 1273-1284.
[30] Shahrabi, J., Hadavandi, E., Asadi, S., Developing a hybrid intelligent model for forecasting problem: Case study of tourism demand time series, Knowledge-Based Systems 43(2013) 112-122.
[31] Takefuji Y, Lee KC, Aiso H, An artificial maximum neural network: a winner-take-all neuron model forcing the state of system in a solution domain, Biol Cybern 67(1992) 243-251.
[32] A.I.Vakhutinsky, B.L.Golden, A hierarchical strategy for solving traveling salesman pronlems using elastic nets, Journal of Heuristics 1(1995) 67-76.
[33] Durbin.R., D. Willshaw, An analogue approach to the travelling salesman problem using an Elastic Net approach, Nature 326(1987) 689-691.
[34] Durbin.R., R.Szeliski, A.Yuille, An analysis of the Elastic Net approach to the travelling salesman problem. Neural computation 1(1989) 348-358.
[35] Yuki Todo, Hiroki Tamura, Kazuya Yamashita, Zheng Tang, Unsupervised learnable neuron model with nonlinear interaction on dendrites, Neural Network 60(2014) 96-103.
[36] Chen, C., Lai, M., Yeh, C., Forecasting tourism demand based on empirical mode decomposition and nueral network, Knowledge – Based Systems 26(2012) 281-287.
[37] Kuan-Yu Chen, Combining linear and nonlinear model in forecasting tourism demand, Expert Systems with Applications 38(2011) 10368-10376.
[38] Bates, J.M., Granger, C.W.J., Combination of forecasts, Operations Research Quarterly20 (1969) 451-468.
[39] Chen, K., Combining linear and nonlinear model in forecasting tourism demand, Expert System with Applications 38(2011) 10368-10376
[40] Law, R., Demand for hotel spending by visitors to Hong Kong: A study of various forecasting techniques, Journal of Hospitality and Leisure Marketing 6(2000) 17-29.
[41] Lin, C., Chen, H., Lee, T., forecasting tourism demand using time series. Artificial neural networks and multivariate adaptive regression splines:

Evidence from Taiwan, International Journal of Business Administration 2(2011) 14-24.
[42] Lin, K., Pai, P., Lu, Y., Chang, P., Revenue forecasting using a least-squares support vector regression model in a fuzzy environment, Information Sciences 220(2013) 196-209.
[43] Yuki Kawakubo, Tatsuya Kubokawa, Modified conditional AIC in linear mixed models, Journal of Multivariate Analysis 129(2014) 44-56.
[44] Xin Li, Bing Pan, Rob Law, Xiankai Huang, Forecasting tourism demand with composite search index, Tourism management 59(2017) 57-66.
[45] Xin Xu, Rob Law, Wei Chen, Lin Tang, Forecasting tourism demand by extracting fuzzy Takagi --- Sugeno rules from trained SVM_s, CAAI Transactions on Intelligence Technology 1(2016) 30-42.
[46] Cho, V., A comparison of three different approaches to tourist arrival forecasting, Tourism management 25(2003) 323-330.
[47] Chu, F.L., Forecasting tourism demand: A cubic polynomial approach, Tourism Management 25(2004) 209-218.
[48] Claveria, O., Torra, S., Forecasting tourism demand to Catalonia: Neural networks vs. time series models, Economic Modelling 36(2014)220-228.
[49] Ping-Feng Pai, Kuo-Chen Hung, Kuo-Ping Lin, Tourism demand forecasting using novel hybrid system, Expert Systems with Applications 41(2014) 3691-3702.
[50] Qu, H., Zhang, H.Q., Projecting international tourist arrivals in East Asia and the Pacific to the year 2005, Journal of Travel Research 35(1996) 27-34.
[51] Shen, S., Li, G., Song, H.,Combination forecasts of international tourism demand, Annals of Tourism Research 38(2011) 72-89.
[52] http://admin.taiwan.net.tw/statistics/year.aspx?no=134 Tourism Burean M.O.T.C. Republic of China (Taiwan).
[53] http://www.mlit.go.jp/kankocho/siryou/toukei/shukuhakutoukei.html Japan Tourism Agency Ministry of Land, Infrastructure, Transport and Tourism.
[54] http://www.jnto.go.jp/jpn/statistics/data_info_listing/index.html Japan National Tourism Organization
[55] http://fdc.fang.com/index/BaiChengIndex.html China
[56] http://industry.fang.com/ China