Research on GM-LSTM Hybrid Model for Tourism Prediction Based on One Belt and One Road

by

Shuxin Zheng

A dissertation
submitted to the Graduate School of Science and Engineering for Education
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Engineering

University of Toyama
Gofuku 3190, Toyama-shi, Toyama 930-8555 Japan

2020
(Submitted June 22, 2020)

Acknowledgements

At frist, I want to thank the people who cared and helped me during my Ph.D course. Every progress in these three years cannot be separated from their help and support. Particularly, I should be very grateful to Prof. Zheng Tang, who has given a lot of guidance and suggestions on the topic selection, topic opening, content and structure of the paper. When I meet difficulties, He always give me encouragement and support to help me overcome them. High tribute should be given to Prof. Shangce Gao, who has put forward to making suggestions for further revisions and devote his a lot of time to guide my thesis title selection and research work. It is very difficult for me to complete my doctoral dissertation without his genuine effort and dedication in polishing and revising my drafts. And I must thank the all members of Tang Lab for sharing their research findings and experience and helping everywhere. Finally, I am sincerely grateful to my family. Without my parental help to take care of my boy who did his best to keep me from worrying during my Ph.D. and my husband's support and encouragement, it is difficult for me to finish my study in peace.

Abstract

The value and significance of the ancient road of silk are fully analyzed, and the cultural connotation and the role of promoting economic development of One Belt and One Road proposed by China are summarized in this thesis. Due to the rich tourism resources along the silk road: high grade, good quality, complete types, strong attraction, with the characteristics of "cultural heritage and natural scenery concurrence , ancient culture and modern civilization coexist". The integration of culture and tourism has strengthened exchanges and mutual learning among different cultures and promoted the integration of cultural and tourism resources around the world. With the support of modern transportation, a unique tourism economic belt has been formed. Therefore, to promote and jointly build the high-quality One Belt and One Road through tourism development, will further break regional and industrial barriers, further promote inter-civilization as well as economic and cultural exchanges and strengthen cooperation in technological innovation, infrastructure and transportation construction, and establish an international division of labor system featuring mutual trust, rationality and win-win results, and make industrial structures complementary and mutually beneficial, and break new ground of regional cooperation and the integration of the economy in Asia and Europe. Therefore, the development of One Belt and One Road tourism has great market potential and great economic and cultural value.

Along with the development in the tourism industry circle , the study of tourism and management put forward higher and more precise requirements. Artificial neural network is a pure computing model based on human brain tissue structure, succeeded in solving many practical problems in the modern computer, and it also has a outstanding performance in the study of the tourism problems, provides a good theoretical basis of the management and decision-making of the tourism. Since tourism data are in a small data set with observation data discreteness and nonlinearity, there are big limitations to prediction mode.

The following models are introduced in this thesis. The grey model, utilizing differential equations to characterize the complex system and making short-term prediction; the rolling window scheme for grey model; increasing the forecasting accuracy; LSTM, solving the convergence problems faced by traditional neural networks for the time sequence forecast, a kind of recurrent neural network (RNN); RBFNN model, Two new hybrid nonlinear dynamic prediction methods based on the neural network of radial basis function and iterative nonlinear filter; ARIMA model, well-establishing time series model for tourist forecasting; and GM-LSTM hybrid model, integrating the first-order grey model and LSTM neural network with a rolling mechanism. So two groups of models with good performance in tourism prediction are selected: the models related to neural network and classical traditional models. For convenience of comparison, three models are selected from each group:

RBFNN with IEKF and IUKF, LSTM and SDNM for the models related to neural network, and classic traditional models including ARIMA, GM and Rolling GM. Then, the same data set and the same simulation software are used for calculation to predict the arrivals to Xi'an from the countries along the OBOR. After the evaluation results, the best performance models of each group was screened out. Then, it is compared with the GM - LSTM model proposed in this thesis.

On the basis of studying the classical model and the neural network model, this thesis uses the mobile window technology combined with GM to obtain rolling - GM, which is used to predict the development trend of the data. Then, the LSTM model is used to predict the residual, and the predicted values of the above two are added up to obtain the predicted results.

Among the above models, the GM-LSTM hybrid model is proposed in this thesis. We take the prediction of the arrival number of tourists from countries along the route of One Belt and One Road to Xi 'an as the target and use the above models for analysis and calculation. Then the results of comparison of the working models show that the the GM-LSTM hybrid model integrates the advantages of the grey model (GM) and the neural network (NN), and can make self-adaptive prediction based on the observation of small samples. The first-order grey model is applied to describe the overall trend of tourism demand, and the nonlinear residual fluctuation of tourism demand is described by the LSTM network with rolling mechanism. Based on the case analysis of the annual arrival number of international tourists to Xi 'an from 1980 to 2018, the hybrid model of GM-LSTM is evaluated in effectiveness and compared with the standard time series model in result. The study result shows that the proposed hybrid model of GM-LSTM is more accurate and more effective for tourism prediction. Therefore, the GM-LSTM model is a more effective model for time series prediction with a small amount of data and high degree of nonlinearity. The model can not only predict tourism problems, but also has a good reference value in the prediction of other similar events.

Contents

Acknowledgements **ii**

Abstract **iii**

1 Introduction **1**

1.1 The study background of tourism 1

1.2 The significance of tourism predictions 7

1.3 The main research content and innovation points of this thesis 8

2 Research on tourism prediction **10**

2.1 Domestic research status about tourism. 12

2.1.1 Research on algorithms and modeling related to tourism problems. 13

2.1.2 Domestic research on One Belt and One Road tourism. 13

2.2 Research status abroad 15

3 Prediction and models **19**

3.1 Prediction 19

3.1.1 Time series 19

3.1.2 Filtering process 20

3.1.3 Prediction classification 20

3.2 Models 20

3.2.1 The models related to neural network 21

3.2.1.1 SDNM (Single Dendritic Neuron Model) 21

3.2.1.2 The radial basis function neural network(RBFNN) 23

3.2.1.3 LSTM introduction (Long Short-Term Memory) 27

3.2.2 Traditional prediction model 32

3.2.2.1 Grey model 32

3.2.2.2 Rolling window selection for gray model 33

3.2.2.3 ARIMA model 35

3.2.3 GM-LSTM model with rolling windows 38

3.2.3.1 Method framework 38

3.2.3.2 GM-LSTM model 39

3.3 Construction of credibility on tourism model: the case of data mining in tourism . . 41

4 Experiments **45**

4.1 Dataset description . . . 45

4.2 Evaluation . . . 45

4.3 Model implementation . . . 46

4.3.1 The models related to neural network implementation . . . 46

4.3.1.1 SDNM implementation . . . 46

4.3.1.2 RBFNN implementation . . . 47

4.3.1.3 LSTM model implementation . . . 48

4.3.2 Traditional prediction model implementation . . . 49

4.3.2.1 Grey model . . . 49

4.3.2.2 Rolling model . . . 50

4.3.3 ARIMA model . . . 51

4.3.4 GM-LSTM model . . . 52

5 Results and discussion **54**

5.1 The results comparison of models related to neural network model . . . 54

5.1.1 The results comparison of the RBFNN models with IEKF and IUKF . . . 54

5.1.2 The results comparison of models related to neural network: SDNM, LSTM and IEKF-RBFNN . . . 55

5.2 The results comparison of traditional models . . . 56

5.2.1 The results of ARIMA . . . 56

5.2.2 The results of GM and Rolling GM . . . 60

5.2.2.1 The results of GM . . . 60

5.2.2.2 The results of rolling GM . . . 62

5.2.3 The results comparison of models: ARIMA, GM and the rolling GM . . . 63

5.3 GM-LSTM model . . . 66

5.3.1 The results of models . . . 66

5.3.2 The result comparison of LSTM model, rolling GM model and GM-LSTM model 68

5.4 Quantitative analysis of three models . . . 72

6 Conclusion **75**

7 Outlook and plan **77**

Bibliography **79**

List of Figures

1.1 The figure of economic belt of OBOR . . . 3
1.2 The comprehensive contribution of tourism economy to China's GDP . . . 4

3.1 Time series . . . 19
3.2 Filtering process . . . 20
3.3 The architecture of SDNM . . . 22
3.4 Radial basis function network structure . . . 24
3.5 The architecture of RNN . . . 27
3.6 The architecture of LSTM . . . 28
3.7 The process of judgment and calculation of LSTM . . . 29
3.8 The operating principle of valve . . . 29
3.9 The architecture of scroll window . . . 35
3.10 Overview framework of hybrid GM-LSTM model . . . 38
3.11 LSTM model of time series prediction (A) and structure of LSTM neuron (B) . . . 39
3.12 Data mining process . . . 42
3.13 Reliability analysis of tourism online evaluation . . . 44

4.1 The schematic diagram of the Rolling model . . . 51

5.1 The standardized residual . . . 58
5.2 The histogram plus estimated density . . . 58
5.3 The plot of residual quantile distribution . . . 58
5.4 The plot of residual autocorrelation . . . 59
5.5 The fitting plot of ARIMA model . . . 59
5.6 The residual error plot of ARIMA model . . . 60
5.7 Comparison of the calculated results of the GM model with the original data . . . 61
5.8 The residual error plot of GM model . . . 61
5.9 The length of rolling window and corresponding mean absolute percentage error (MAPE) of one-step-ahead prediction of rolling $GM(1, 1)$ on the training set . . . 62
5.10 The comparison of the results: the rolling GM model and the original data . . . 63
5.11 The residual error plot of rolling GM model . . . 64
5.12 Comparison of $ARIMA(2, 2, 4)$ and $GM(1, 1)$ on the training and test set . . . 64
5.13 Comparison of rolling $GM(1, 1)$ and $GM(1, 1)$ on the training and test set. . . . 65
5.14 The residual error plot of rolling GM model and GM model . . . 66

5.15 Comparison of the calculated results of the rolling LSTM model with the original data 67
5.16 The residual error plot of LSTM model 67
5.17 Comparison of the calculated results of the GM-LSTM model with the original data . 68
5.18 The residual error plot of GM-LSTM model 69
5.19 Comparison of the calculated results of the GM-LSTM model with the original data . 69
5.20 Comparison of the GM-LSTM, rolling GM, and LSTM model with the original data . 70
5.21 Comparison of the calculated results of the errors 72

List of Tables

1.1 Percentage of selected items (multiple choices) 6

4.1 International tourist arrivals in Xi'an from 1980-2018 (Unit: ten thousand) 46

5.1 The forecasting results from 2014-2018 (Unit: ten thousand) 55
5.2 Comparisons of AMSE for yearly tourist arrival forecasting 55
5.3 Results of SDNM, LSTM and IEKF-RBFNN 55
5.4 The calculation results of the three models 57
5.5 The calculating results of the models of rolling window-GM, ELSTM and the GM-LSTM 71
5.6 The calculating results of the models of MSE, MAE and MAPE 73

Chapter 1

Introduction

In recent years, the Chinese industrial structures have been adjusted and changed greatly in the coexisting period of industrialization and post-industrialization due to the all over the world economic slow growth and the big economic downward pressure, and the tertiary industry has been greatly developed. In 2013, the GDP of China's tertiary industry exceeded that of its secondary industry for the first time. Tourism industry, as an important component of the tertiary industry supported by the government and with sustainable development of regional ecology and economy, is rising like a shining star. However, cultures play an very essential role in tourism. In some instances, cultures are the best resources for tourism, and tourism is the biggest market for cultures. Guidelines on promoting integrated development of cultural tourism jointly issued by the ministry of the national tourism administration and culture in September 2009, put forward that culture is the main soul of the tourism and the tourism is an very important carrier of the culture. In the process of developing the tourism industry, the integration of culture and tourism provides strong impetus for the development of the new tourism formats. The core of the new tourism formats is the integration of culture and tourism. Tourism and culture support and promote each other for their developments. While greatly enriching the tourism products, the integration of culture and tourism also develops the cultural products and promotes and drives the growth of the cultural industry, and it has become the driving force in the development of the tourism industry and the cultural industry, as well as an effective measure to realize the 'Site Activation', cultural utilization, ecological and economic sustainable development.

1.1 The study background of tourism

The integration of Cultural and tourism is in a new period of industrial development. In the process of development, cross-border cultural exchanges are also promoting key cooperation in global tourism [1].

In September 2013, China proposed to build the silk road economic belt with nearly 3 billion population, strengthen exchange with all kinds of civilization, promote peace and world development, and drive economic prosperity and regional cooperation among countries along the belt route, and to work with together the ASEAN countries to build a 21st century maritime silk road. The economic belt of silk road and the maritime silk Road are called as One Belt and One Road (OBOR or Belt and

Road for short). On June 22, 2014, the eastern section of the overland silk road named the Routes Network of Chang'an-Tianshan Corridor' jointly declared by China, Kyrgyzstan and Kazakhstan was successfully declared as a world cultural heritage site, and became the first transnational cooperation item of the world heritage list. The new silk road economic belt, with the Asia-Pacific economic circle in the east and developed European economic circle in the west, covers about 4.4 billion people and has an economic aggregate of about 21 trillion us dollars, accounting for 63% of the world's population and 29% of the world's economy, respectively. It is considered as "the world's longest and most promising economic corridor", and the silk road economic belt has a vast territory, and is abundant in natural resources, mineral resources, energy resources, land resources and the precious tourism resources. It is well known as the strategic energy and resource base of the 21st century. In March 2015, China issued the document "vision and actions for jointly building the silk road economic belt and the 21st century maritime silk road", and the OBOR strategy was put into action. The construction of OBOR has ushered in a new opportunity for the tourism industry of the whole countries, including China, and the whole regions along the silk road , and has provided a great development space for China's cultural and tourism integration innovation.

The silk road which carries the unique history and culture was first proposed by German geographer Ferdinand von Lichthofen in 1877. The silk road is an ancient commercial trade route that began in the political, economic and cultural center of ancient China, Chang 'an (now Xi 'an), and connects the ancient trade routes of Asia, Africa and Europe. It crosses the Longshan mountains, and the Hexi corridor, through the Yumen pass and Yangguan pass to Xinjiang and then finally to Africa and Europe along the Pamirs through the Centre of Asia, West Asia and North of Africa, and it was a major road for economic, political and cultural communications between east and west, and the important trade route by which China output a lot of silk, porcelain, spices and other products to the world. Take silk, for example, it is said that silk was invented by Leizu in ancient times, and silk has been passed down in China for thousands of years, so China is the hometown of silk. Silk, as a luxury product, because of its strong practicability, high value and non-perishable characteristics, was often used as a settlement when money was scarce, and became a common international currency at that time. Silk, which has been passed down for thousands of years, carried and spread China's unique history and culture to every corner of the world through this historic international trade channel, and made essential contributions to the economic, political and cultural communications between the east and the west in ancient China. Therefore, this trade channel, silk road, is known as the world's most important commercial artery [2].

The construction of OBOR has brought new opportunities to the tourism of China, even the whole countries and the whole regions along route of the silk road, and has provided a great development space for China's cultural and tourism integration innovation. As a concentrated embodiment of the soft power for a country, culture and tourism play a special role and strategic position in the overall competitiveness of a country [3]. The cultural and tourism markets in the OBOR region are extremely important not only to China but also to the whole world. The OBOR line contains extremely rich cultural tourism resources, has a huge market volume and the corresponding high-speed industrial growth, the integration of culture and tourism development will inject vitality into the development of tourism market along the line [4].

The special history and culture endowed the ancient silk road with uniqueness and irreplaceability, and laid a good historical and cultural foundation for the development of OBOR today due to the historical value and cultural value of the silk road.

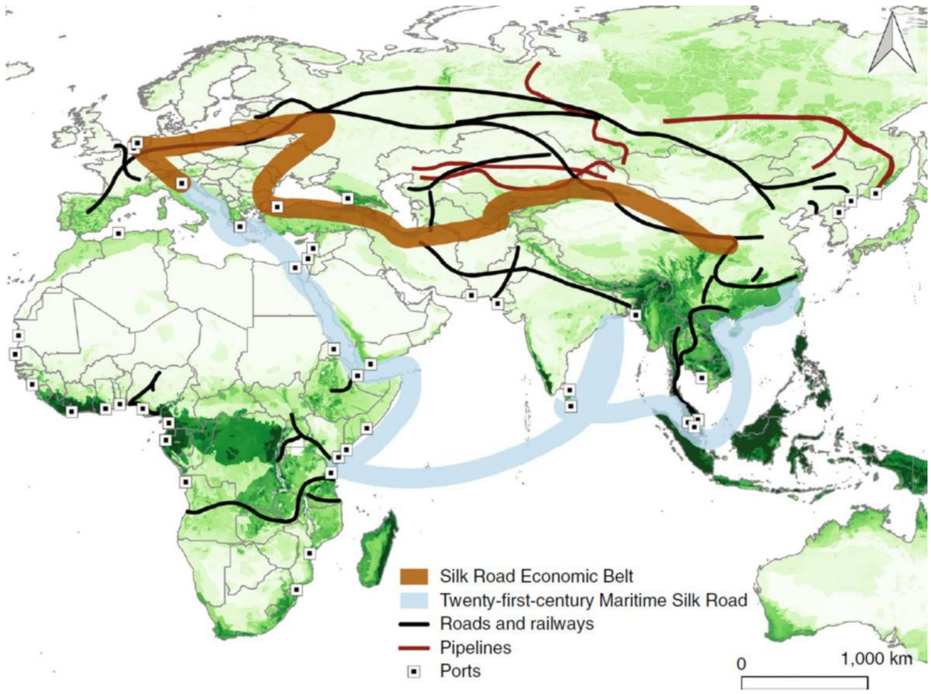

Figure 1.1: The figure of economic belt of OBOR

The maritime silk road east sea route was formed during the Qin and Han dynasties, that is, around 200 BC. The original sea route followed the coastline of the Liaodong peninsula and the Korean peninsula to the south and then crossed the Tsushima strait to the north of the Japanese archipelago. In the Sui and Tang dynasties, with the joint efforts of Chinese and Japanese people, there were many sea routes across the yellow sea and the east China sea. Through these sea routes, not only Chinese goods were continuously exported to Japan and the Korean peninsula, but Chinese culture was also spread to these countries on a large scale, including Confucianism, legal system, Chinese characters, clothing, architecture, even martial arts, tea drinking customs, etc. The maritime silk route the south China sea route was also formed during the Qin and Han dynasties, as evidenced by the discovery of African ivory and other imports from the Nanyue kingdom (203-111 BC) in the pearl river delta region. In 111 BC, Emperor Wudi of the Han dynasty sent his army to subjugate the state of Nanyue and directly control the gate of the south China sea shipping route. With his powerful national strength, Emperor Wudi of the Han dynasty, with great talent and great strategy, greatly expanded overseas transportation and opened the first ocean route directly from the south coast of China to the Indian Ocean, reaching as far as the east coast of the present Indian peninsula and Sri Lanka. This route is clearly recorded in the book of geography of Han dynasty. At the western end of Eurasia, the Roman empire was also expanding rapidly. After the fall of the Ptolemaic dynasty in Egypt in 30 BC, the Romans began to push eastward through the red sea into the Indian Ocean. In the first century, the Romans mastered the Indian Ocean's monsoon patterns and used them to sail directly across the Indian Ocean, to and from the red sea to India, rather than just along the Indian Ocean's coastline as before. Some traders from the Roman empire even sailed to the southeast coast of the Indian peninsula to establish trading posts. The account of the voyage of the Eritrean sea, a Greek nautical document formed in the second half of the first century AD [5].

With the development of trade, the huge capital flow promoted the prosperity of merchants along the silk road. Silk alone consumed half of all coins made in the Roman empire each year. The nomadic tribes of the moon established large commercial centers for all kinds of goods from India, central Asia, and China, boosting the economies of the countries between China and Rome.

In history, the silk road was not only a paradise for goods and a holy land for the creator, but also a holy palace for wisdom exchange. In the early 4th century, Emperor Constantine built a splendid new capital on the site of Byzantium on both sides of the Bosporus, which became a bridge between the Mediterranean and the eastern world. Various religious sects with different political colors also successively gathered on the gold-strewn silk road. They were decisive on the battlefield or at the negotiation table, and even influenced the rise and fall of a country. The silk road in history has influenced the global economy and even the world's political structure.

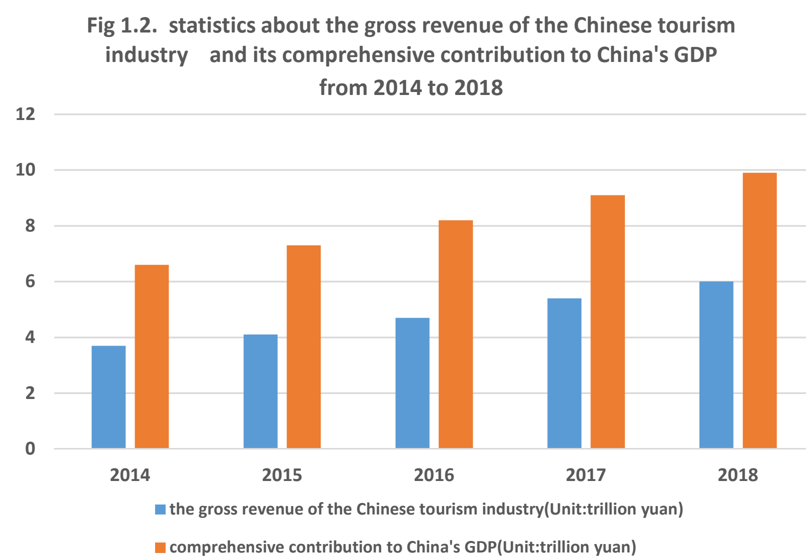

Figure 1.2: The comprehensive contribution of tourism economy to China's GDP

In the early 21st century, trade and investment on the ancient silk road came alive again. European and central Asian countries hope to expand cooperation with China. Under the background of rapid development of modern transportation, information and globalization, promoting the development and cooperation in various fields of economy and trade along the silk road is not only the inheritance of history and culture, but also the development of the huge potential of the region [6].

Inheriting the open tradition of the ancient silk road, OBOR strategy upholds the spirit of openness and inclusiveness and promotes the development of countries along the route of OBOR. The silk road economic belt runs from Asian countries all the way to European countries, forming a trend of common economic development in Eurasia. The strategy of the 21st century maritime silk road economic belt connects the three continents of Asia, Africa and Europe and the economic belt of the silk road by sea, forming an economic ring that circulates endlessly at sea and on land.

The OBOR is once again like a solid long line connecting countries, economy, culture, religion, art and transportation along the line. One Belt and One Road is more like a container, integrating countries, economy, culture, religion and art along the line and influencing each other. Today's resumption of the silk road will bring new opportunities, cooperation, challenges and success.

On the basis of the ancient silk road, OBOR not only attaches great importance to economic and cultural exchanges, but also strengthens cooperation in technological innovation, infrastructure and transportation construction. It also establishes an international division of labor system featuring mutual trust, rationality and win-win, makes the industrial structure complementary and mutually beneficial, and creates a new pattern of regional cooperation and economic integration in Asia and Europe.

Under the background of OBOR, the development of transportation and the blending and complementation of culture are gradually influencing the unbalanced development of the export and import of tourism culture at home and abroad, and changing the surplus and deficit of tourism trade. OBOR is the path of the ancient and modern civilization, China, as silk road starting and reviving the silk road proposer, connects the ancient civilizations and modernization of countries along the route, connects the world's largest international tourism market and the silk road, and integrates the development of silk road and tourism, fully plays to the huge market potential, and greatly promotes the development of global tourism culture. According to a report by I-media showed in Fig 1.2, from 2014 to 2018, the total tourism industry revenue of China, and its comprehensive contribution to China's GDP all increases year by year by the end of 2018, at the same time, the annual contribution of tourism revenue to the comprehensive contribution to China's GDP is also increasing year by year, which is 56.1%, 56.2%, 57.3%, 59.3% and 60.6% respectively. For a instance, the annual contribution of tourism revenue reaches 6.0 trillion yuan RMB in 2018, and its comprehensive contribution to China's GDP is 9.9 trillion yuan RMB, accounting for 60.6% of the comprehensive contribution GDP, and it indicates tourism has become a new economic growth point of the nation, and the tourism status is becoming more important in economic development. The development of tourism has provided a new opportunity and platform for the transformation from the primary industry to the secondary industry and from the secondary industry to the tertiary industry. The establishment of this new industrial chain and the clear division of labor have boosted the economic growth of all walks of life, improved the overall production network, and played a very essential role in the smooth transformation of China industrial structure.

The development of tourism promotes the international currency circulation and increases the national foreign exchange income. Tourism foreign exchange income is significant to economic growth, and it exists a two-way causal relationship [7].

Interaction between tourism and OBOR, not only brings economic benefits , but also promotes the various levels of cultural exchange between countries, eases the communication barriers formed by the factors such as history, geography and culture, increases the understanding to make the cooperation of countries along the more solid, more forceful, promotes the world peace and spreads culture well [8].

The region along the silk road is rich in tourism resources: high grade, good quality, all types, strong attraction, coupled with the support of modern transportation routes, forming a unique tourism economic belt. The characteristics of its tourism resources are: the coexistence of cultural heritage and natural scenery, and the coexistence of ancient culture and modern civilization. So the silk road tourism has the strong conditions to appeal to tourists all over the world.

Since 1978, the national tourism administration has always insisted on the silk road as the country's gold tourism line, and the silk road tourism is one of world-class tourism products recommended by the world tourism and organization one of the products that are popular in the international tourism market. Since the 1990 s, the silk road project received the guidance and support by both of the United Nations cultural, scientific and educational organization and the world organization of tourism [9].

China is one of the four ancient civilizations. The most representative image perception element of foreign tourists is the history and culture of ancient China. According to the survey of CNN, for activities they can experience in China, the proportions of international audiences choosing historical sites, natural landscape, art and culture, and traditional festivals are 89.50%, 86.81%, 82.82% and 78.74%, respectively. On the other hand, the tourism terms of their interest in China, the proportions of international audiences.

Table 1.1: Percentage of selected items (multiple choices)

	historical sites	natural landscape	art and culture	traditional festivals
experience	89.50%	86.81%	82.82%	78.74%
interest	75.72%	69.73%	59.88%	53.68%

To choose historical sites, natural landscape, art and culture, and traditional festivals are 75.72%, 69.73%, 59.88% and 53.68%, respectively. It shows hat foreign tourists have great interest in China's historical sites and culture [10].

As the most representative of China's historical and cultural destination and the core destination of inbound tourism, Xi 'an is not only the ancient capital of the 13 dynasties, such as Zhou, Qin, Han, Tang dynasties, but also the first of the four ancient capitals all over the world. Xi 'an is home to two thousand years of culture, including the world-famous terracotta warriors and horses. At the same time, the world cultural heritage project: in the Chang'an-Tianshan corridor network, Xi 'an occupies five heritage sites including: the Greater Wild Goose Pagoda, built in 652 to preserve the Sutra and Buddha statue that Xuanzang monk brought back to Chang 'an from India via the silk road by which spread the buddhist culture; founded in 707, Lesser Wild Goose Pagoda built to preserve the buddhist scriptures brought back by Yijing is at the Jianfu Temple where it is located is one of the three major sutra translation fields in Chang 'an in the Tang dynasty, which proves the history of Buddhism's spread from India to the east, and also witnesses the popularity of Buddhism in Chang 'an in the Tang dynasty; located in the Xingjiao temple in Xi 'an city, the Xingjiao temple pagoda bears witness to Xuanzang's co-translation of buddhist scriptures and the development of Buddhism in east Asia; built in 634, the ruins of Daming Palace in Chang 'an city of the Tang dynasty, is a representative site of the oriental starting point during the heyday of the silk road in the 7th to 10th centuries; built in 200 BC in the Han dynasty, the ruins of Weiyang palace is the earliest eastern silk road starting point .

Xi 'an, as the ancient silk road starting point and the starting point of today's OBOR, carries the spirit and the core value endowed by history and the new era. Tourism, as a pillar industry of culture in the new era, must be deeply integrated with technology and science, so as to continuously improve the scale and professional level of the tourism industry, and enable the ancient civilization to shine with the support of modern science and technology. Xi 'an, as one of the first batch of national demonstration bases of the integration of technology and science and culture , has a good base for the information industry and modern science and technology culture foundation, at the same time, Xi 'an, as today's international metropolis, was named the capital of culture in east Asia and with the continuous activation of the traditional culture and historical sites, as well as the modern information technology's depth excavation to the city's tourism resources, will be in a whole new look, new customer experience and profound cultural deposits to attract visitors from all over the world.

Although Xi 'an is located in the inland, transportation conditions have been greatly improved during the One Belt and One Road's construction , which has greatly improved the accessibility of tourism from the coast to the inland and the border tourism between neighboring countries and China's border provinces. Therefore, this study chooses xi 'an as an example.

1.2 The significance of tourism predictions

With the booming tourism market of cities along the OBOR line, the number of inbound tourists and domestic tourists is increasing. While tourism brings benefits to cities along the line, it also brings great challenges to the reception capacity, service capacity and environmental safety index of scenic spots, and has a certain impact on the lives of local residents. This thesis forecasts the number of inbound arrival of tourist in a city, which is helpful to the policy makers' overall planning, rationally developing, and coordinating of tourism resources. By predicting the change of tourist flow, we can make scientific and reasonable planning for the construction and configuration of transportation and public facilities, make reasonable adjustments to the tourist routes and travel plans, accurately orient tourists' needs and tourism products, and make reasonable planning and adjustment to the structure of tourism products, improve the service quality of scenic spots, enhance sense of participation and experience. To control the flow of visitors, the city and scenic spots can be effectively managed and channelized to avoid accidents. Strengthen the linkage of various scenic spots, so that traffic, hotels, scenic spots and shopping achieve seamless docking, drive the development of the whole industry chain, create a new model of cultural tourism, improve the attention and reputation of tourist cities. In an age of the big data and the artificial intelligence , the information revolution's influence on society is enormous, using the big data and the artificial intelligence to forecast the entry, into the number of scenic spots, helps policy makers to coordinate development of the tourism market and limited ecological environment, the relationship between the sustainable development of the scenic spot between the environment and resources, long-term stability of the local tourism planning. To sum up, the prediction of urban passenger flow and scenic spot passenger flow is very necessary for the planning of cities and scenic spots, as well as an essential part of urban development. At the same time, in the process of development in tourism transmission culture, the effective protection of historical and cultural heritage should be realized, so that it can be better inherited and play a greater role in promoting the continuable development of tourism and enhancing the core competitiveness of tourism products.

On another hand, One Belt and One Road tourism strategy is an irreplaceable role in promoting the economic development of countries along One Belt and One Road, and the exchange of world culture, strengthening the international influence of Chinese civilization, accelerating the export of Chinese civilization and culture, and realizing the strategy of regional development and common prosperity. We will promote high-quality development of One Belt and One Road, promote exchanges and mutual learning among civilizations, and further promote people-to-people connectivity.

In the new era of integrated development of culture and tourism, the culture and tourism system "Promote tourism through culture and highlight culture through tourism" continues to promote cultural exchanges and cooperation by exploring the cultural connotation of Chinese culture, especially the 'OBOR' initiative. Exchanges and cooperation in the fields of culture and tourism will help further break down regional and industrial barriers, integrate the world's cultural and tourism resources, and further consolidate the consensus and social foundation for the 'OBOR' , and the building of a community of 'shared future for mankind'.

Therefore, the research on One Belt and One Road tourism has great economic value and historical and cultural significance.

1.3 The main research content and innovation points of this thesis

Through the above review of research on demand of tourism under the background of One Belt and One Road and the review of research on tourism-related issues by big data technology, It is found that the research on tourism under the background of One Belt and One Road has the following deficiencies: First, the research on One Belt and One Road tourism is more qualitative and less quantitative, lacking scientific theoretical research. Second, previous studies on this issue are limited to the analysis of the impact of tourism and the tourism demand of one country or one region under the 'OBOR' strategy. Thirdly, the data span of previous studies is small and the data volume is insufficient. Fourth, It is not found that the big data technology is used to analyze the problems of tourism market prediction under the background of 'OBOR'.

Therefore, the main contents and innovations of this thesis are as follows:

This thesis presents a hybrid model of GM-LSTM. We take the prediction of the number of tourists to Xi 'an as the target and use the above models for analysis and calculation. The comparison results of the working models show that the hybrid model integrates the advantages of the grey model and the neural network, and can make adaptive prediction for small samples. The first-order grey model is used to describe the overall trend of tourism demand, and the LSTM network is applied to describe the nonlinear residual fluctuation of tourism data. By analyzing the number of international tourists in xi 'an from 1980 to 2018, the mixed model is evaluated, and the result of computation are compared with the standard time series model. The result shows that the mixed model is more accurate and more effective for tourism prediction.

Chapter 2

Research on tourism prediction

From the perspective of tourists, tourism usually can be divided into seven.

Sightseeing tourists: the main purpose of sightseeing tourists is to appreciate the places of interest, local conditions and customs in a foreign country. At the same time, it can also be combined with shopping, entertainment, investigation and official business. It is the oldest, most common and most fundamental type of tourists in the world and the main body of Chinese tourists. It is characterized by hoping to enjoy the natural landscape and cultural landscape of foreign countries, increase the knowledge, broaden the vision, edify the sentiment, get new, strange, different, beautiful, special feelings; Short stay in the tourist destination, low revisit rate, less cost, more sensitive to the characteristics of tourist attractions and prices.

Recreational tourists: recreational tourists aim to relax the spirit, enjoy the temporary change brought by the new environment. As recreational tourism can regulate the pace of people's life, get rid of the daily tension caused by the task of the annoyance, this type of tourists are increasing. Of all the tourists in the developing countries, recreational tourists account for the largest proportion. Its characteristics are the pursuit of entertainment, participation, recreation, excitement and enjoyment; with sensitive the quality, safety, price of tourism products; Travel season is strong, it almost will choose the best travel destination season, and the use of paid vacation travel; The choice of tourist destination and travel mode has great freedom; The revisit rate is higher, and the travel and stay time are longer.

Official tourists: official tourists are tourists who, according to the requirements of tasks, take trade cooperation, business negotiation, attending meetings, holding exhibitions, scientific and cultural exchanges as the main purpose, and run the activities of visiting and sightseeing under the premise of completing official duties. Its characteristic is that it has the certain status to the tourism product and the service quality request is high; The expenses are mainly the public expenses of the organization, and the ability to pay is strong, not sensitive to the price, and the consumption is high; Because of official business , the travel destination and travel time do not have too much choice, generally with the nearby short distance and short time more; The number of people is relatively small, but the travel frequency is more, the seasonality is not strong.

Personal and family tourists: the needs of these travelers are complex. They differ in their needs from those of leisure and business, but they have some of the characteristics of both. For example, in terms of travel time, although many of them use paid holidays to visit relatives and friends, but quite a lot of people choose to go out to visit relatives on traditional holidays, and the traditional holidays are not uniform in various countries. In addition, many family and personal affairs, such as attending a wedding, attending the opening ceremony and other date restrictions are tight. Therefore, its overall characteristics are that it is sensitive to travel price; There is no freedom to choose a tourist destination.

Medical health tourists: medical health tourism mainly includes health tourism, leisure tourism, hot spring tourism, forest tourism, sports health tourism, specialized Qigong tourism and other forms. The main purpose of health care tourists is to cure some chronic diseases and eliminate the fatigue of daily tasks through participating in the tourism activities beneficial to physical and mental health Its characteristic are that the tourists have higher income, more free time, strong desire to stay healthy or recover; It is sensitive to the functions of health care, recreation and health care in tourism projects. The proportion of middle and old people is larger and the residence time is longer.

Cultural knowledge-oriented tourists: cultural knowledge-oriented tourism is a type of tourism that aims to see the society, experience the folk customs, accumulate rich historical culture and increase knowledge. The main purpose of cultural knowledge tourists is to achieve active rest and entertainment through cultural knowledge tourism, and to obtain enlightenment and enrichment of knowledge. Its characteristic are the high cultural accomplishment, the strong thirst for knowledge, have a certain specialty or special interest, willing to discuss with others; There are higher requirements on the cultural knowledge of tour guides, and more sensitive to the rigor of the travel schedule and the scientific nature of the tour routes.

Eco-tourists, also called adventure traveler: adventure tourism is a kind of new advanced tourism emerging in the international and domestic tourism market. It emphasizes the combination of sightseeing tourism, nature conservation and cultural conservation, and is a new tourism project with environmental responsibility and environmental ethics. The main purpose of adventure tourists is to get in touch with nature, understand nature, publicize and protect the nature of the active rest and entertainment.

Since the main research object of this study is tourists to Xi 'an, and the study aims at the cultural tourism, the tourists are mainly composed of cultural knowledge tourists, sightseeing tourists and recreational tourists

There is a lot of literature on the analysis of these three types of international tourists, especially the use of artificial intelligence and big data technology to research and predict the number of arrivals, the geographical distribution of the number of tourists and the cultural attraction of the research has achieved good results. The following is an overview of domestic and foreign research on tourism.

With the development of the whole world economy, the global tourism is also growing rapidly. At the same time, there are great changes that have taken place in the tourism industry: the continuous expansion of the industry scale has led to the explosive growth of the information and data of tourism. Tourism data has a huge information space to grow , making it possible to analyze and predict the tourism market with big data technology. It has become a new direction of tourism for big data application how to make full use of the massive original tourism data and to mine and analyze the massive travel data accumulated daily, and reflecting tourist information, quickly, accurately and conveniently.

Big data is a data set composed of a large number with complex structure and many types of data. It is an intelligence resource and knowledge service capability formed by data integration, sharing and cross-utilization based on data processing and application mode [11-12].

Data mining, also known as knowledge data discovery, means that extracts valuable information hidden in data from large, fuzzy, incomplete, random and noisy data. It is one of the hottest topics in the field of the database research and the artificial intelligence at present. It is a support decision process. Based on machine learning, pattern recognition, database, artificial intelligence, etc., it analyzes a large amount of data automatically, and makes inductive reasoning, and mines potential patterns to provide decision-making support for the departments of government. Data mining techniques mainly include sequence pattern, association analysis, classification, anomaly detection, clustering and so on. For the big data application of tourism, correlation analysis can be used to study the data of tourism, and find out the high probability patterns, or through data clustering and classification, the similarity of tourism data can be analyzed and similar data can be stored together to provide support decision for decision-makers. Big data analysis can play an irreplaceable role in mining valuable tourism information, mining potential tourism customers, optimizing tourism route strategies, recommending tourism projects and destinations, and especially in predicting the tourism market.

On the research of tourism big data analysis, predecessors have done a lot of work.

2.1 Domestic research status about tourism.

Guo xin introduced the concept of big data and the demand of big data in tourism development, and summarized the common technology of the mining technology of tourism big data and data mining methods, which can realize the mining of valuable tourism information, mining potential tourism customers, optimizing tourism routes and recommending tourism projects and destinations. It is proposed that tourism big data technology will obtain more valuable information and more market opportunities [13]. Lu yuan analyzed the intelligent tourism mode from the perspective of big data. Through the establishment of database, the use of big data, government promotion, etc., this paper analyzed the overall hierarchy diagram of intelligent tourism, and put forward three levels: integration layer, basic processing of original data; data model: the network-based data analysis and mining model is established through the data model, and the three-dimensional model is finally formed to ensure that different data can be generated in various dimensions to provide data support for tourism industry development, the presentation layer: mainly based on the data model, displays the existing data, provides graphic data for users and serves the whole industry [14]. Zhang xiaohua etc. , analyzed the technical problems and difficulties to be solved in the smart tourism software., and the information processing model for the core problems of the industry is studied, so that users only need to input their own travel plan, they can get rid of the tedious travel planning and immediately get a complete, efficient, feasible and optimized travel plan [15]. Tian qiuyang applied big data analysis in the tourism industry to classify consumers, and based on this, mined the specific needs of users, accurately positioned the tourist destinations and consumer groups of customers, which could improve the economic benefits of enterprises [16]. Aiming at the development trend of big data in local tourism, Xiao jie analyzed and elaborated the application points of big data in tourism management. On the basis of information data collection, operation and management, as well as big data analysis, a high-quality processing mode can be created to ensure the personalized positioning of the brand market, and on this premise, the inter-industry acceptance and public recognition can be improved. At the same time, the establishment of a large database of consumer groups can help them grasp the contents of consumption behaviors and interests and preferences by means of statistics and analysis, so as to grasp the product reputation in the market [17]. Zhang lijun and Zhao xia deployed the cloud platform based on cloud computing technology for massive base station data storage, introduced big data analysis and combined GIS map, by analyzing the user's behavior of staying in the scenic spot and so on, accurately modeling, to provide users with tourism information and services, at the same time, the scenic spot management staff effectively grasp the situation of the scenic spot personnel flow, targeted management. [18]

2.1.1 Research on algorithms and modeling related to tourism problems.

Liu xiaoyan, using artificial intelligence technology, a tourism management information system of precise destination marketing is proposed, which is modeled based on actual tourism market survey data, and intelligently screens various behaviors of the model by using knowledge base, so as to provide the tourism destination and decision-making for marketers accurately. The results indicate that the algorithm can use the neural network technology to provide automatic classification, and through experiments, the management information system based on the algorithm provides accurate data support for tourism marketing. [19]

Meng zhihui found that the travel time of tourists presented a very irregular state, is is caused by the asymmetry of information acquisition between the two sides. Therefore, it is proposed to use wi-fi coverage, data information integration and data cleaning and other methods to master the changes of scenic spot passenger flow [20].

Sun yanping proposed an artificial neural network based tourist source prediction method, which can be used to predict tourist source. It is an effective method to predict tourist sources effectively, and the result is better than the traditional method, with high accuracy and robustness, and strong adaptability to the model [21].

Using artificial intelligence method, Yu mingtao optimized the BP neural network with the modified particle group algorithm, and improved the optimization performance of the particle group through the non-linear recursion of the inertia factor. This prediction model is applied to the prediction of tourist flow of scenic spots, and the accuracy of prediction is improved through the experimental simulation of multiple training samples and test samples with few parameters, high effective.Effective research has been carried out in predicting passenger flow [22].

2.1.2 Domestic research on One Belt and One Road tourism.

Jing le thought our country can only be called tourism country instead of great tourism power, in terms of service trade development scale and market structure, put forward with the stratagem of Belt and Road to bring new opportunities for the development of tourism service trade, promote the reformation of the tourism consumption structure, strengthen the international competitiveness in tourism industry circle, and strengthen the application of the information technology [23]. Based on the background of One Belt and One Road, Cheng qian expounded the new pattern and new development of China's tourism service trade. Finally, countermeasures to develop tourism service trade under the background of One Belt and One Road were proposed [24]. Wang zhanlong made a overall analysis on the impact of One Belt and One Road on tourism, so as to further optimize the space and pattern of tourism development, so as to lay a solid basis for the continuable and rapid development of the tourism industry in China [25]. Han zhiyong analyzed the influence of One Belt and One Road on ASEAN tourism. Tourism is an industry that ASEAN countries focus on developing. The development of ASEAN tourism can not be apart from China, and the development of China's tourism cannot be separated from ASEAN countries. Under the new situation, with the proposal of One Belt and One Road and the implementation of relevant policies, new challenges and opportunities have been brought to the cooperative development of China and ASEAN's tourism [26]. Taking silk road tourism as an example, Song hongjuan etc. used data mining technology to judge the market value of tourism and predicted its destination selection behavior. They mainly used three data mining procedures: RFM analysis (recent travel time, travel frequency and travel cost) to identify valuable silk road travelers; The decision tree is used to analyze the social attribute characteristics, the decision-making behavior pattern and the choice of tourist destination of the valuable tourists. Basket analysis for travel destination preferences for cross-selling [27]. Previous researchers used RFM to analyze real transaction records or obtained data through questionnaires, but few studies were conducted through data mining [28,29]. Most of the studies on silk road tourism remained at the macro level, while few were conducted at the micro level. This paper systematically applied big data technology to analyze tourism demand and value judgment, but did not predict the number of people in a certain destination, nor did it analyze the influencing factors of tourism revenue.

Yang ting proposed that One Belt and One Road brings together the essence of the world's tourism resources and can promote cultural and economic exchanges between Xi 'an and countries and cities along the land and sea silk roads. At present, we can take advantage of the opportunity of development in silk road tourism to meet international standards, improve the product quality and service quality of China's tourism products, promote the tourism innovation in China's system of management, and accelerate the steps of "going global" of China's traditional culture and Xi 'an's tourism industry [30]. Su hongxia et al. took the inbound number of international tourists in Shaanxi province from 2007 to 2016 as the sample, and analyzed the spatio-temporal evolution characteristics of inbound international tourists in Shaanxi province by spatial concentration index model, landscape-attraction model and competitive state model. The results show that the international tourist market of Shaanxi is mainly dominated by developed countries in America and Europe , while the countries along the route of the belt and the road account for a small share of the market, mainly in southeast Asia, and many countries along with the route of silk road from central Asia to eastern Europe have not been prominent in the international market of Shaanxi [31]. This paper analyzes the characteristics of the dynamic evolution of Shaanxi's tourist sources. The international community has become increasingly recognized and engaged in the One Belt and One Road initiative. One Belt and One Road has brought tangible benefits to countries and regions along with the belt and road in project construction, economic and trade cooperation and other fields. Thanks to extensive and in-depth cultural cooperation and communications with foreign countries, people of all countries have continuously enhanced their cognition and understanding of One Belt and One Road [32].

According to the forecast of the world tourism organization, by the end of 2020, the number of outbound Chinese tourists will exceed 100 million, and Chin will be the world's fourth largest source of tourists. In fact, in 2018, the actual number of outbound Chinese tourists reached 149 million, an increase value by 14.7 percent over that of the same period of last year, and the number of inbound tourists was 141 million, up 1.2% than that of last year. International tourism revenue was $127.1 billion, up 3.0 percent from a year earlier [33].

In the era of globalization in which the Chinese economy and the world economy interact deeply, China's outbound tourism market and consumption, which constantly create new records, will undoubtedly have a profound impact on the world development in tourism pattern. Therefore, it is of great significance to study the tourism theory and application under the One Belt and One Road stratagem.

2.2 Research status abroad

María Henar etc studied three data sources that reflect different travel activities in the city: panorama (sightseeing), Foursquare(consumption), and Twitter(connected accommodation). Several data sources must be used in complementary ways to analyze the digital footprints of city visitors [34]. Sheng-Hshiung Tsaur, etc discussed the application of artificial neural network in loyalty analysis in international tourist hotels and the neural network modeling of loyalty based on business determinants. The results explains that the neural network model has well effect in prediction [35]. Stephen F. Witt and Lindsay W. Turner, use the comprehensive method to forecast the number of international passenger arrivals in the source market and destination region of mainland China, and the time series is combined with the measurement method, which is called the quantitative analysis of the structural comprehensive time series [36]. Alfonso Palmer, etc, introduces the basic principle of artificial neural network and provides a method for neural network design of travel time series prediction step by step. This method has been successfully verified by using the artificial neural network, using the time series of tourism expenditure in the Balearic islands (Spain) as data, and providing reference for researchers interested in applying artificial neural network to the prediction of tourism data [37]. The influence of tourism forecast combination on accuracy is studied in detail by Chi Kin Chan etc., a quality control technology, was introduced for the first time to determine the time of updating weight, and a hybrid method (quadratic programming) was developed to combine prediction to reduce prediction errors. Empirical results manifest that the control weight method not only saves the updating time of combination weight, but also improves the overall performance of combination prediction.This study puts forward a new idea to improve the accuracy of prediction [38]. Since the role of tourism demand prediction in tourism management is very complex and important, Jamal Shahrabi proposed a hybrid model combining the data preprocessing and genetic fuzzy expert system, namely modular genetic fuzzy prediction system (MGFFS), to meet the needs of tourism prediction. In order to study a tool that can predict the development of tourism industry accurately, this system is composed of three parts: the data preprocessing is the first stage . In the time series model, the key lags to be considered are selected by statistical test. On this basis, the k-means clustering methods and data transformation are used to establish a model of a modular to reduce the complexity of the all data space and make it more unified. And the genetic learning algorithm is efficiently adopted to allocate fitness through symbiotic evolution and extract the system from each cluster according to TSK fuzzy rules in the second stage. Finally, the experimental data are clustered and the fuzzy system of each cluster is used to predict the tourism demand.

Results show that the accuracy of the prediction compared with the classical time series model, neural fuzzy system and neural network model of prediction method has great improvement, can be applied to forecast trends and patterns, and points out that the size, direction and future international tourism flow, can travel for the government and private sector organizations strategy and marketing strategy, can also be customized suitable tourism products. The research of this paper has a practical significance and strong theoretical value for the booming tourism industry [39]. Oscar Claveria, Salvador Torra, Combining more advanced prediction techniques with growing interest in tourism need and demand to more accurately predict demand at the level of destination for the growth of the world tourism, it enables us to assess the predictive performance of models of the neural network relative to time series methods at the level of region. The volatility and seasonality are important characteristics of travel data, so comparing the predictive performance of linear models and nonlinear alternatives is a particularly advantageous background. When comparing the prediction accuracy of the autoregressive comprehensive moving average model, the accuracy in prediction of the autoregressive threshold autoregressive model and the model of thr artificial neural network is better than that of the autoregressive model and the artificial neural network model in different time ranges, especially in the short time range. These results show that there is a trade-off between the degree of preprocessing and the accuracy of neural network prediction, and neural network prediction is more suitable for nonlinear data [40]. Yen-Hsien Lee and Ya-Ling Huang Combined with the grey model (GM) of prediction and the Fourier residual correction model, the prediction effect of the random fluctuation data is improved, so that it can estimate the fluctuation in the historical time series. In this study, grey model prediction and Fourier residual correction model are combined to improve the effectiveness of random fluctuation data prediction, improve the accuracy in short-term prediction and the volatility of major case samples. This is very practical for today's international tourism market. The study has great significance to the complexity, variability and uncertainty of China's tourism market under the stratagem of One Belt and One Road [41].

Haiyan Song and Stephen F. Witt used the vector autoregression (VAR) model to predict the passenger flow to Macau from eight major source countries, and the reason for using VAR technology is that it allows impulse response analysis to study how the demand of Macau's tourism industry responds to the "shocks" of the variables of economy in the VAR system. The meaning of the analysis is discussed. The prediction of VAR model shows that Macao will face the growing tourism demand of mainland Chinese residents. As the demands of Chinese tourists are often different from those of other countries of origin, especially in the West, the business sector in Macau needs to attach great importance to meeting the needs of Chinese tourists [42]. Chen wei proposed a new model of SDNM (single dendritic neuron model) to predict travel demand. The characteristics of tourism were analyzed by using phase space reconstruction technique, and the time series was reconstructed into suitable phase space points. Then, the maximum lyapunov index was used to identify the chaotic characteristics of time series and to determine the prediction limit. Finally, the model of SDNM was used for short-term prediction. The experimental results of Japan's monthly inbound tourist forecast show that this method has higher efficiency and interaction than neural networks such as multi-layer perceptron, neuro-fuzzy reasoning system, Hermann network and single-neuron model [43]. Different forms of autoregressive moving average models (ARMA) are the most widely used in time series prediction, and they have also been widely applied to tourism prediction and have achieved excellent results [44]. Jungmittag [45] suggested a seasonal autoregressive integrated moving average (SARIMA) model for prediction of monthly air travel demand with obvious seasonal patterns at the German airport Frankfurt am Main. Liang [46] applied the SARIMA and the generalized autoregressive conditional heteroskedastic model to analyze and predict the tourism need in Taiwan and compared with other forecasting models. Ma et al. [47] used the ARIMA model to calculate 36-month-ahead forecast and compared the forecast errors of eight different competing ARIMA models. Recently Li et al. presented an effective tourist flow forecasting model based on Long Short-term Memory (LSTM) the neural networks [48]. However, these data-driven ANNs often require a large dataset for training. The performance would be limited if only small-sample observations are provided.

There are a lot of foreign articles analyzing tourism demand by using big data technology, with rich research contents and comprehensive research problems. It focuses on using artificial neural networks to predict travel destinations, accommodation, air tickets, or arrivals for a particular country. However, the prediction of tourist arrivals under the background of One Belt and One Road is basically absent, while the research on the prediction of tourist arrivals in xi 'an under the background of One Belt and One Road has not been found. At the same time, there are few theoretical researches on tourism under the background of One Belt and One Road, and even fewer practical researches on tourism market.

Through researching the literature on One Belt and One Road tourism, it is found that there are many articles on the significance study of One Belt and One Road tourism, while there are fewer theoretical study or quantitative study on One Belt and One Road tourism. About the theoretical research papers, there are more analyses on the evolution of market and tourism space, but less analyses using big data technology or artificial intelligence, and there are many Chinese literature and few foreign literature for the research on OBOR.

In summary, the current research on tourism prediction has the following characteristics: (a) in terms of the categories of models, more models related to neural networks are used, while less traditional ones are used; There are more single models and fewer composite models; (b) for tourism prediction with less data and high degree of nonlinearity, there are few studies on the validity of the model and the prediction accuracy should be improved; (c) the study on tourism in One Belt and One Road has more Chinese literature and less foreign literature (almost none); In the Chinese literature, there are more qualitative researches and less quantitative researches on One Belt and One Road tourism; (d) there is less research literature on Xi 'an as a tourist destination, and even less research literature on tourism from One Belt and One Road countries along the route to Xi 'an. The quantificational prediction research on the number of tourists from the countries along with the route One Belt and One Road to Xi 'an has not been found.

Therefore, whether it is the spread of ancient eastern civilization or the exchange and integration of world culture; Whether it is the construction of world economic community or the economic development of the countries along with the route of One Belt and One Road , the research of this paper has practical application significance and theoretical research value.

Chapter 3

Prediction and models

3.1 Prediction

Time series prediction analysis is to use the past a period of time a certain event time characteristics to forecast future event of a period of time of the event characteristics. It is a kind of relatively complex forecast modeling problem, for this kind of problem in general, there are two types of models: the model of regression analysis and the model of the neural network , time series model of the neural network is dependent on the order of the events, his input variable is a set of time sequence of number sequences, and the same size value is the result of a change order after the input model is different, so the difficulty is bigger than regression forecasting. Prediction should be made up of management and making - decision activities and play an crucial important role in many decision making areas. Many modern organizations need to make short-term, medium-term and long-term forecasts according to specific application requirements.

3.1.1 Time series

As to be mentioned above, there are many researches and methods on tourism prediction, but there are few literatures on study deeply of One Belt and One Road tourism prediction. Therefore, on bases in the tourism data of One Belt and One Road, this chapter analyzes the validity of various models for prediction. First of all, this is a prediction of time series, and its basic principle is as follows, and time series is the series of data points arranged in the order of time occurrence. Usually

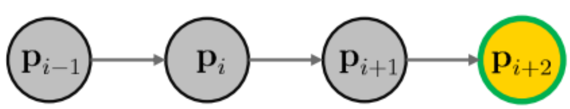

Figure 3.1: Time series

the time interval of a set of the time series is a constant value (such as 1 second, 5 minutes, 12 hours, 7 days, 1 year), so the time series can be used as distance analysis and processing of scattered time data:

$$\mathbf{T} = \{\mathbf{p}_0, \cdots, \mathbf{p}_n\}, \mathbf{p}_i = \{\mathbf{x}_i, t_i\}. \tag{3.1}$$

3.1.2 Filtering process

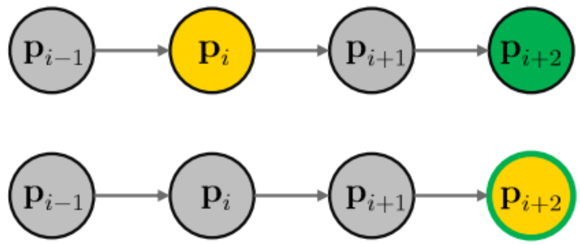

Figure 3.2: Filtering process

Smooth: the current moment is $i+1$, and the moment to be processed is i, which is given according to the data in the window around the data possible values at time i (savitzky-golay Filter);

Denoising: signal is collected at moment $i+2$ and needs to be processed at moment $i+2$. The optimal estimation value at moment $i+2$ is expected to be obtained through processing (Kalman Filter).

3.1.3 Prediction classification

Prediction should be consist of the decision making activities and the management because it has good performance in management and decision making. All modern organizations require long-term, medium and short forecasts, rely on the specific application. The short-term forecasting the scheduling of production, personnel, and transportation requires short-term forecasting. As part of the scheduling process, you often need to anticipate requirements as well. Medium and long-term forecasts require determining future resource requirements to plan for the purchase of raw materials, medium term forecasts requires the hiring of personnel, or the purchase of machinery and equipment. The long-term predicting events are used for strategic - planning and these decisions should take into account the opportunities of markets, internal resources and the environmental factors. Prediction must be to predict the future event in result as accurately as possible, taking into account all available information including the knowledge of any future events that may affect the prediction and historical data . The problem we aim at is short-term forecasting.

3.2 Models

The appropriate forecasting method depends on the available data to a large extent. If the data is not available, or the available data is less relevant to the prediction , the qualitative prediction methods should be used. However these methods could not be pure guess, some well-developed structured methods can yield good predictions without using the data of history. Quantitative prediction can be applied with two conditions, which met the needs: the available digital information on the past; The assuming that some aspects of past patterns will continue into the future is reasonable.

There are many kinds of quantitative forecasting methods, which are usually developed for specific purposes in specific disciplines. Every method has its given characteristic, accuracy, and cost, which should be considered when selecting a particular method. The most quantitative prediction problems use either cross-sectional data (collected at certain time points) or time series data (collected at regular intervals in time). We are more concerned with predicting future data in the time series domain.

The models of tourism problem research can be divided into three categories: the models related to neural network, the traditional classical models and the composite models in prediction. Through the literature, the models related to neural network, and the traditional classical models have good performance of time series forecast, so we select some of the best model in the prediction effect of tourism from the two types of models, and based on the data of tourists entering Xi 'an, analysis the effectiveness of the model and the prediction accuracy. Then try to combine these two kinds of models, generate a new combination model, and to study the same problem, find a more effective prediction model.

3.2.1 The models related to neural network

Artificial neural network is a prediction method based on a simple mathematical model of the brain. They allow for complex nonlinear relationships between response variables and their predictors. The Neural network is a biological neural system simulation based on operations of biological neural network. In recent years, some researches of artificial neural network has made big progress and solved many problems successfully in the automatic control, biology, intelligent robot and the economics areas, and practice in the field of pattern estimation and recognition. Basically, ANN has been shown to have good intelligence and the performance of neural network in tourism prediction is also very good.

3.2.1.1 SDNM (Single Dendritic Neuron Model)

SDNM is made up of four layers including one layer of synaptic performs sigmoid function, the dendritic layer acts as the output synaptic of the multiplication function, the membrane layer is actually the output of an addition function all the dendritic branches, however, another s-shaped function is uses by a soma function to output the entire single neuron results.

Synaptic Layer Synapses can be inhibitory or excitatory, wholly depending on ionization energy caused by the change in postsynaptic potential. The function that connects i-th $(i = 1, 2, \cdots, N)$ synaptic input to the j-th $(i = 1, 2, \cdots, M)$ synaptic layer can be expressed by Eq.(3.1). The value k is a positive value invariant, weight w_{ij} and threshold θ_{ij} the connection parameter.

$$Y_{ij} = \frac{1}{1 + e^{-k(w_{ij}x_i - \theta_{ij})}}. \tag{3.2}$$

When the initial value of k is too large, the sigmoid function will become a step function. Depending on the values of w_{ij} and θ_{ij}, the four connection examples exist: (1) A constant 0 connection (when $w_{ij} < 0 < \theta_{ij}$ or $0 < w_{ij} < \theta_{ij}$). If the input is changed from 0 to 1, the output is

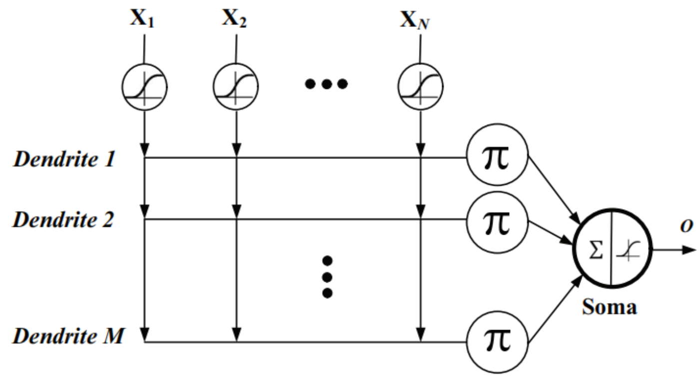

Figure 3.3: The architecture of SDNM

0. (2) A constant 1 connection (when $w_{ij} < 0 < \theta_{ij}$ or $0 < w_{ij} < \theta_{ij}$). If the input value is changed from 0 to 1, the output value will become 0. (3) Excitatory connection (when $0 < \theta_{ij} < w_{ij}$). If the input is changed from 0 to 1, the output equals the input and the synapse will be in an excitatory type. (4) Inhibitory connection (when $w_{ij} < \theta_{ij} < 0$) where the synapse the output will reverse and input will be an inhibitory type in this case.

Dendrite Layer The dendritic layer consists of the nodes among each branch. We must be noted that we use the soft minimization operator to cope with the binary input classification problem in the previous dendritic neuron model, while the multiplication operation used in this study can deal with the real input problem. The equation output of the j-th branch is shown below.

$$Z_j = \prod_{i=1}^{N} Y_{ij}. \tag{3.3}$$

Membrane Function Branching results are summed by an operation that is similar to the logic OR operation in the binary case. The output approximation is as follows.

$$V = \sum_{j=1}^{M} Z_j. \tag{3.4}$$

Soma Function The output can be calculated as follows

$$O = \frac{1}{1 + e^{-k_{soma}(V - \theta_{soma})}}. \tag{3.5}$$

The parameter k_{soma} is set as a positive constant and the threshold θ_{soma} is variable from 0 to 1.

BP-like Learning Method The output should be calculated as follows.

$$E = \frac{1}{2}(T - O)^2. \tag{3.6}$$

w_{ij} and, θ_{ij} are connected in the learning process. The output vector generated by the input vector is The compared with the target vector. It can reduce the error between the output vector and the teaching signal T by correcting the w_{ij} and θ_{ij}. The error value of the calculation method is shown as follows:

$$\Delta w_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}}, \tag{3.7}$$

$$\Delta \theta_{ij}(t) = -\eta \frac{\partial E}{\partial \theta_{ij}}, \tag{3.8}$$

where η is the learning constant and it represents a positive constant [43]. The updating rules for w_{ij} and θ_{ij} are defifined as:

$$w_{ij} = w_{ij} + \Delta w_{ij}(t), \tag{3.9}$$

$$\theta_{ij} = \theta_{ij} + \Delta \theta_{ij}(t), \tag{3.10}$$

where t is the learning epoch and the differentials of E with respect to w_{ij} and θ_{ij} can be expressed as follows

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial V} \cdot \frac{\partial V}{\partial Z_j} \cdot \frac{\partial Z_j}{\partial Y_{ij}} \cdot \frac{\partial Y_{ij}}{\partial w_{ij}}, \tag{3.11}$$

$$\frac{\partial E}{\partial \theta_{ij}} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial V} \cdot \frac{\partial V}{\partial Z_j} \cdot \frac{\partial Z_j}{\partial Y_{ij}} \cdot \frac{\partial Y_{ij}}{\partial \theta_{ij}}. \tag{3.12}$$

The application of SDNM model in prediction will be discussed in the following chapters.

3.2.1.2 The radial basis function neural network(RBFNN)

A RBFNN is composed of m-dimensional input $u = \begin{bmatrix} I_1 & I_2 & \cdots & I_m \end{bmatrix}$ passed directly to a hidden layer with c neurons. Each neuron in the hidden layer applies an activation function that is a function of the Euclidean distance between the m-dimensional prototype vector and the input. Then the output of the neuron of the hidden layer is weighted and transferred to the layer of output. The output of RBFNN is composed of the weighted sum of hidden neurons layer. Fig. 3.4 indicates a schematic of RBFNN and the response of it can be written as (where the function in hidden layer has the form of $g(s) = \exp[-s/\beta^2]$ and β is a real constant):

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \omega_{11} & \cdots & \omega_{1c} \end{bmatrix} \begin{bmatrix} g\left\|u - v_1\right\|^2 \\ \vdots \\ g\left\|u - v_c\right\|^2 \end{bmatrix} = \omega \begin{bmatrix} g\left\|u - v_1\right\|^2 \\ \vdots \\ g\left\|u - v_c\right\|^2 \end{bmatrix}. \tag{3.13}$$

The nonlinear filter based RBFNN training model Generally, We can regard the optimization problem of weight matrix as the weighted least squares problem and the prototype problem as the weighted least squares problem, Where the error vector is the difference between the target values of these outputs and the output of RBFNN. Consider the RBFNN of Fig.3.4 with n outputs, c prototypes and m inputs. We let the elements of the elements of the prototypes v_j and the weight

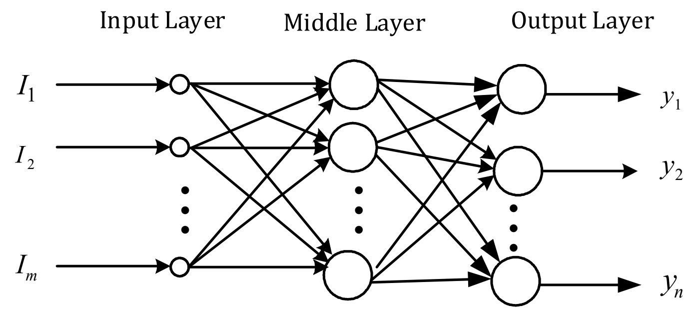

Figure 3.4: Radial basis function network structure

matrix ω constitute the state vector of a nonlinear system, and let the outputs of the RBFNN constitute the outputs of the nonlinear system to which the nonlinear filter is used in order to cast the problem in optimization in a form suited for the nonlinear filtering methods. And the state in the nonlinear system can be expressed as:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{\omega}^T & \boldsymbol{v}_1^T & \cdots & \boldsymbol{v}_c^T \end{bmatrix}^T, \tag{3.14}$$

The vector $\boldsymbol{x}$ includes all parameters of the RBFNN for $(n(c+1)+mc)$ arranged in a linear array. We need to add some artificial process noises $\boldsymbol{w}_k$ and observation noise $\boldsymbol{v}_k$ to the system model in order to execute a stable nonlinear filtering algorithm. The model of a nonlinear system to which the nonlinear filter can be shown is

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{v}_k, \quad \boldsymbol{y}_k = h\left(\boldsymbol{x}_k, \boldsymbol{u}_k\right) + \boldsymbol{w}_k. \tag{3.15}$$

where $\boldsymbol{x}_k$ is the state vector in the system at time k, $\boldsymbol{y}_k$ is the observation vector, and $h(\bullet)$ is the nonlinear vector function for the input vector and the state vector.

Iterative nonlinear filters The general nonlinear filters such as UKF and EKF are all based on the principle of the minimum mean squared error (MSE) , and the main difference between the the iterative nonlinear filters and general nonlinear filters lies in the step update of measurement [48].

The iterated extended Kalman filter Mean iterative update equations of IEKF and its covariance with approximation of Gaussian to obtain the posterior distribution of its state are given,

1. Initialized with $\hat{\boldsymbol{x}}_0 = \mathrm{E}\left[\boldsymbol{x}_0\right], \widehat{\boldsymbol{P}}_0 = \mathrm{E}\left[\left(\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_0\right)\left(\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_0\right)^T\right]$.

2. For $k = 1, 2, \cdots$.

   (a) Time update

$$\bar{x}_k = f\left(\widehat{\boldsymbol{x}}_{k-1}\right), \overline{\boldsymbol{P}}_k = \boldsymbol{F}_k \widehat{\boldsymbol{P}}_{k-1} \boldsymbol{F}_k^T + \boldsymbol{Q}_k. \tag{3.16}$$

   where $\boldsymbol{Q}_k$ is the process noise covariance and $\boldsymbol{F}_k$ is the Jacobian matrix of the state equation of transition.

   (b) Measurement update equations

After obtaining the corresponding covariance $\overline{\boldsymbol{P}}_k$ and the $\bar{\boldsymbol{x}}_k$, the iteration will be carried out recursively.

$$\bar{x}_{k,0} = \bar{x}_k, \qquad \overline{\boldsymbol{P}}_{k,0} = \overline{\boldsymbol{P}}_k, \tag{3.17}$$

$$\boldsymbol{H}_{k,j} = \frac{\partial h}{\partial x}|\boldsymbol{x} = \overline{\boldsymbol{x}}_{k,j}, \qquad \boldsymbol{K}_{k,j} = \overline{\boldsymbol{P}}_k \boldsymbol{H}_{k,j}^T \left[\boldsymbol{H}_{k,j}\overline{\boldsymbol{P}}_k\boldsymbol{H}_{k,j}^T + \boldsymbol{R}_k\right]^{-1}, \tag{3.18}$$

$$\widehat{\boldsymbol{x}}_{k,j} = \overline{\boldsymbol{x}}_k + \boldsymbol{K}_{k,j}\left[\boldsymbol{y}_k - h\left(\overline{\boldsymbol{x}}_{k,j}, \boldsymbol{u}_k\right)\right], \widehat{\boldsymbol{P}}_{k,j} = \overline{\boldsymbol{P}}_k - \boldsymbol{K}_{k,j}\boldsymbol{H}_{k,j}\overline{\boldsymbol{P}}_k, \tag{3.19}$$

where $\boldsymbol{R}_k$ is the covariance matrix for measurement noise, and j is the iterative number of the same measurement value. The iterative process will be continuing until a certain condition of termination is met. For iterated number $j = 1, 2, \cdots, N$, the corresponding covariance and the ultimate outputs of the state estimation matrix are

$$\widehat{\boldsymbol{x}}_k = \widehat{\boldsymbol{x}}_{k,N}, \quad \widehat{\boldsymbol{P}}_k = \widehat{\boldsymbol{P}}_{k,N}. \tag{3.20}$$

The iterated unscented Kalman filtering Similar to the IEKF, in order to obtain the state posterior distribution, the IUKF also needs to update the mean and the covariance using the method of Gaussian approximation, and its main steps are given below.

1. Initialized with $\hat{\boldsymbol{x}}_0 = \mathrm{E}\left[\overline{\boldsymbol{x}}_0\right], \widehat{\boldsymbol{P}}_0 = \mathrm{E}\left[\left(\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_0\right)\left(\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_0\right)^T\right]$.

2. For $k = 1, 2, \cdots$,

   (a) Calculating the associated weights and the sigma points.

$$\begin{aligned}
&\widehat{x}_{k-1,i} = \widehat{\boldsymbol{x}}_{k-1}, \widehat{\boldsymbol{\chi}}_{k-1,i} = \widehat{\boldsymbol{x}}_{k-1} + \left(\sqrt{\left(n_{\boldsymbol{x}} + \lambda\right)\widehat{\boldsymbol{P}}_{k-1}}\right)_i \quad \left(i = 1, \cdots, n_{\boldsymbol{x}}\right), \\
&\widehat{x}_{k-1,i} = \widehat{\boldsymbol{x}}_{k-1} - \left(\sqrt{\left(n_{\boldsymbol{x}} + \lambda\right)\widehat{\boldsymbol{P}}_{k-1}}\right)_i \quad \left(i = n_{\boldsymbol{x}} + 1, \cdots, 2n_{\boldsymbol{x}}\right),
\end{aligned} \tag{3.21}$$

$$\begin{aligned}
&W_0^{(m)} = \frac{\lambda}{\left(n_{\boldsymbol{x}} + \lambda\right)}, \quad W_0^{(c)} = \frac{\lambda}{\left(n_{\boldsymbol{x}} + \lambda\right)} + \left(1 - \alpha^2 + \beta\right), \\
&W_i^{(m)} = W_i^{(c)} = \frac{1}{\left\{2\left(n_{\boldsymbol{x}} + \lambda\right)\right\} \quad i = 1, \cdots, 2n_{\boldsymbol{x}}}.
\end{aligned} \tag{3.22}$$

   where λ is the parameter of the scaling which is defined as $\lambda = \alpha^2(n_{\boldsymbol{x}} + \kappa) - n_{\boldsymbol{x}}$ ($n_{\boldsymbol{x}}$ is the state vector dimension, and the constants κ, β and α are used as adjustment parameters).

   (b) Time updating.

$$\overline{\boldsymbol{\chi}}_{k,i} = f\left(\widehat{\boldsymbol{\chi}}_{k-1,i}\right), \quad \overline{x}_k = \sum_{i=0}^{2n_x} W_i^{(m)}\overline{\boldsymbol{\chi}}_{k,i}, \tag{3.23}$$

$$\overline{\boldsymbol{P}}_k = \sum_{i=0}^{2n_x} W_i^{(c)}\left[\overline{\boldsymbol{\chi}}_{k,i} - \overline{\boldsymbol{x}}_k\right]\left[\overline{\boldsymbol{\chi}}_{k,i} - \overline{\boldsymbol{x}}_k\right]^T. \tag{3.24}$$

(c) Measurement updating.

$$\overline{\boldsymbol{\gamma}}_{k,i} = h\left(\overline{\boldsymbol{\chi}}_{k,i}, \boldsymbol{u}_k\right), \quad \overline{\boldsymbol{y}}_k = \sum_{i=0}^{2n_{\boldsymbol{x}}} W_i^{(m)} \overline{\boldsymbol{\gamma}}_{k,i}, \tag{3.25}$$

$$\overline{\boldsymbol{P}}_k^{yy} = \sum_{i=0}^{2n_{\boldsymbol{x}}} W_i^{(c)} \left[\overline{\boldsymbol{\gamma}}_{k,i} - \overline{\boldsymbol{y}}_k\right] \left[\overline{\boldsymbol{\gamma}}_{k,i} - \overline{\boldsymbol{y}}_k\right]^T, \tag{3.26}$$

$$\overline{\boldsymbol{P}}_k^{xy} = \sum_{i=0}^{2n_{\boldsymbol{x}}} W_i^{(c)} \left[\overline{\boldsymbol{\chi}}_{k,i} - \overline{\boldsymbol{x}}_k\right] \left[\overline{\boldsymbol{\gamma}}_{k,i} - \overline{\boldsymbol{y}}_k\right]^T, \tag{3.27}$$

$$\boldsymbol{K}_k = \overline{\boldsymbol{P}}_k^{xy} / \overline{\boldsymbol{P}}_k^{yy}, \tag{3.28}$$

$$\widehat{\boldsymbol{x}}_k = \overline{\boldsymbol{x}}_k + \boldsymbol{K}_k \left(\boldsymbol{y}_k - \overline{\boldsymbol{y}}_k\right), \tag{3.29}$$

$$\widehat{\boldsymbol{P}}_k = \overline{\boldsymbol{P}}_k - \boldsymbol{K}_k \overline{\boldsymbol{P}}_k^{xy} \boldsymbol{K}_k^T. \tag{3.30}$$

where $\boldsymbol{K}_k$ is the Kalman gain.

i. Let $\hat{\boldsymbol{x}}_{k,0} = \overline{\boldsymbol{x}}_k, \widehat{\boldsymbol{P}}_{k,0} = \overline{\boldsymbol{P}}_k$ and $\hat{\boldsymbol{x}}_{k,1} = \hat{\boldsymbol{x}}_k, \widehat{\boldsymbol{P}}_{k,1} = \widehat{\boldsymbol{P}}_k$. Also let $j = 2$ (where j is the j-th iteration number).

ii. Generate new sigma points

$$\begin{aligned} &\boldsymbol{\chi}_{k,0,j} = \widehat{\boldsymbol{x}}_{k,j-1}, \quad \boldsymbol{\chi}_{k,i,j} = \widehat{\boldsymbol{x}}_{k,j-1} + \sqrt{\left(n_{\boldsymbol{x}} + \lambda\right) \widehat{\boldsymbol{P}}_{k,j-1}} \quad (i = 1, \cdots, n_{\boldsymbol{x}}), \\ &\boldsymbol{\chi}_{k,i,j} = \hat{\boldsymbol{x}}_{k,j-1} - \sqrt{\left(n_{\boldsymbol{x}} + \lambda\right) \widehat{\boldsymbol{P}}_{k,j-1}} \quad (i = n_x + 1, \cdots, 2n_x). \end{aligned} \tag{3.31}$$

iii. Time update and measurement update

$$\bar{\boldsymbol{x}}_{k,j} = \sum_{i=0}^{2n_{\boldsymbol{x}}} w_i^{(m)} \boldsymbol{\chi}_{k,i,j}, \tag{3.32}$$

$$\boldsymbol{\gamma}_{k,i,j} = h\left(\boldsymbol{\chi}_{k,i,j}, \boldsymbol{u}_k\right), \quad \overline{\boldsymbol{y}}_{k,j} = \sum_{i=0}^{2n_{\boldsymbol{x}}} w_i^{(m)} \boldsymbol{\gamma}_{k,i,j}, \tag{3.33}$$

$$\overline{\boldsymbol{P}}_{k,j}^{yy} = \sum_{i=0}^{2n_x} w_i^{(c)} \left[\boldsymbol{\gamma}_{k,i,j} - \overline{\boldsymbol{y}}_{k,j}\right] \left[\boldsymbol{\gamma}_{k,i,j} - \overline{\boldsymbol{y}}_{k,j}\right]^T + \boldsymbol{R}_k, \tag{3.34}$$

$$\overline{\boldsymbol{P}}_{k,j}^{xy} = \sum_{i=0}^{2n_{\boldsymbol{x}}} w_i^{(c)} \left[\boldsymbol{\chi}_{k,i,j} - \overline{\boldsymbol{x}}_{k,j}\right] \left[\boldsymbol{\gamma}_{k,i,j} - \overline{\boldsymbol{y}}_{k,j}\right]^T, \tag{3.35}$$

$$\boldsymbol{K}_{k,j} = \overline{\boldsymbol{P}}_{k,j}^{xy} / \overline{\boldsymbol{P}}_{k,j}^{yy}, \tag{3.36}$$

$$\widehat{\boldsymbol{x}}_{k,j} = \overline{\boldsymbol{x}}_{k,j} + \boldsymbol{K}_{k,j} \left(\boldsymbol{y}_k - \overline{\boldsymbol{y}}_{k,j}\right), \quad \widehat{\boldsymbol{P}}_{k,j} = \widehat{\boldsymbol{P}}_{k,j-1} - \boldsymbol{K}_{k,j} \overline{\boldsymbol{P}}_{k,j}^{yy} \boldsymbol{K}_{k,j}^T. \tag{3.37}$$

where $\boldsymbol{\gamma}_{k,i,j}$ is the i-th component of $\boldsymbol{\gamma}_{k,j}$.

iv. Define equations $\hat{\boldsymbol{y}}_{k,j} = h\left(\hat{\boldsymbol{x}}_{k,j}, \boldsymbol{u}_k\right), \widetilde{\boldsymbol{x}}_{k,j} = \hat{\boldsymbol{x}}_{k,j} - \hat{\boldsymbol{x}}_{k,j-1}$ and $\tilde{\boldsymbol{y}}_{k,j} = \boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k,j}$

v. If $j < N$, then set $j = j + 1$ and return to 2(c)ii; otherwise, continue to 2(c)vi.

vi. Stop if $j \geq N$ and final results of state estimation and corresponding covariance matrix are

$$\widehat{\boldsymbol{x}}_{k|k} = \widehat{\boldsymbol{x}}_{k|k,N}, \quad \boldsymbol{P}_{k|k} = \boldsymbol{P}_{k|k,N}. \tag{3.38}$$

3.2.1.3 LSTM introduction (Long Short-Term Memory)

LSTM is the time recursive neural network that appears to solve a fatal flaw of RNN. The native RNN will encounter a big problem, called the vanishing gradient problem for RNNs, that is, the 'forgetting' phenomenon will appear in the nodes of later time. In order to solve the problem, recursive neural network was used to model the temporal relationship. As a result, LSTM network model was generated and was more effective than traditional RNNS. The circular neural network based on LSTM can solve the problem of multiple input variables almost perfectly. Cyclic neural network based on LSTM can be well used in time series prediction, because many classical linear methods are difficult to adapt to multivariable or multi-input prediction problems, so the LSTM model of network will be widely applied to the time series prediction.

RNN model The most commonly used and powerful tool of time series model is a kind of recurrent neural network (RNN). Compared with the results of ordinary neural network, all results of each hidden layer of RNN are related to the current input and the results of the last hidden layer. By this method, the calculated results of RNN have the characteristics of remembering the previous results.

The typical RNN network structure is as follows:

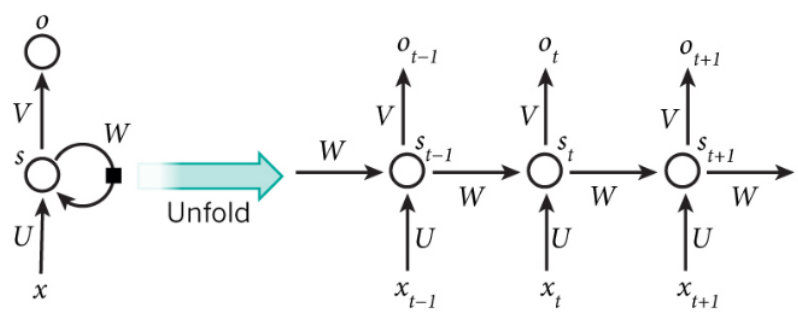

Figure 3.5: The architecture of RNN

On the right is a structure that is easy to understand and memorize when calculating. In short, x is the input layer, and O is the output layer, s is the hidden layer, and t refers to the number of calculations; U, V, W are the weights, where $s_t = f\left(Ux_t + Ws_{t-1}\right)$ is used to calculate the state of the hidden layer at the t-th time, so as to realize the purpose of linking the current input result with the previous calculation.

Because if the RNN model needs to realize long-term memory, it demands to link the calculation of the current hidden state with the calculation of the previous n times, that is, $s_t = f(Ux_t + W_1 s_{t-1} + W_2 s_{t-2} + \cdots + W_n s_{t-n})$, in that case, the calculation amount will exponentially increase, leading in a significant increase for the time of model training. Therefore, RNN model is generally directly used for long-term memory calculation.

LSTM model LSTM (Long Term Memory) model is a variant of RNN, which was proposed by Juergen Schmidhuber firstly. The structure of the classical LSTM model is as follows:

LSTM is characterized by the addition of various layers of valve nodes in addition to the structure of RNN. There are three types of valves: output gate, input gate and forget gate. These valves,

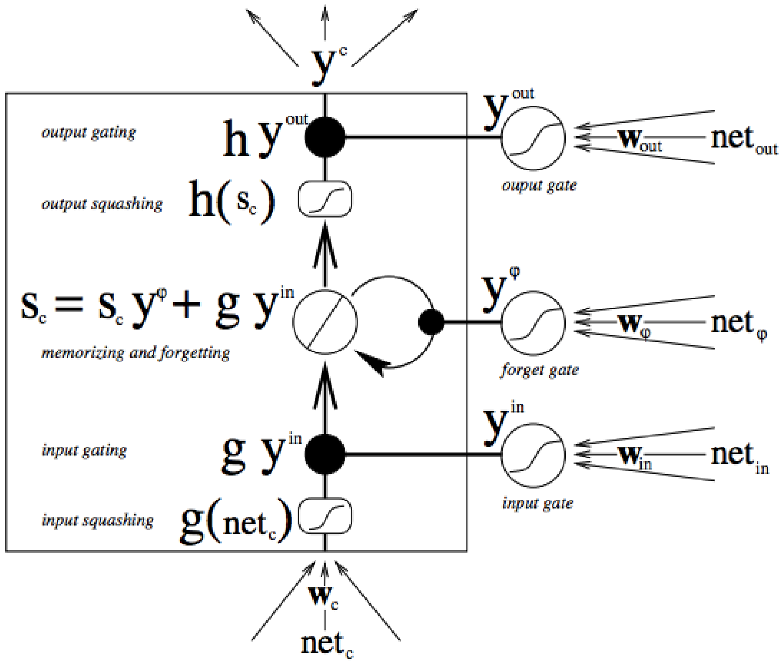

Figure 3.6: The architecture of LSTM

which can be opened or closed, will be used to determine if the memory state of the model network (the state of the previous network) has reached the threshold output in this layer and thus be added to the current layer calculation. As shown in the figure 3.4 and figure 3.6, the valve node uses the sigmoid function to calculate the network state of the memory as the input. If the output reaches the threshold value, the valve output is multiplied with the current layer calculation of result as the input of the next layer (PS: here the multiplication is the multiplication of each element in the pointing matrix); If the threshold is not reached, the output is forgotten. The weight of each layer including the valve node is updated during each model back propagation training. And the more specific judgment and calculation process of LSTM is shown in the following figure 3.7:

The memory function of LSTM model is realized by these valve nodes. While the valve is open, the training results of the previous model will be related to the current model calculation, while the previous calculation results will no longer affect the current calculation while the valve is closed. Therefore, by adjusting the switch of the valve, there can be achieved the effect of the early sequence on the final result. And when you don't want the previous results to have an impact on the next, such as the beginning of a new paragraph or chapter in natural language processing, turn off the valve.

The black solid circle represents the calculation result of the node output to the next layer or the next calculation; The hollow circle indicates that the calculated result of the node has not been input into the network or received a signal from the last time.

The hidden state $h_{(t-1)}$ and the sequence data $x^{(t)}$ are input in the figure. The output $f^{(t)}$ of the forgetting gate is obtained by an activation function, usually sigmoid. As a result of the sigmoid output $f^{(t)}$ between the $[0, 1]$, so the output here $f^{(t)}$ It represents the probability of forgetting the

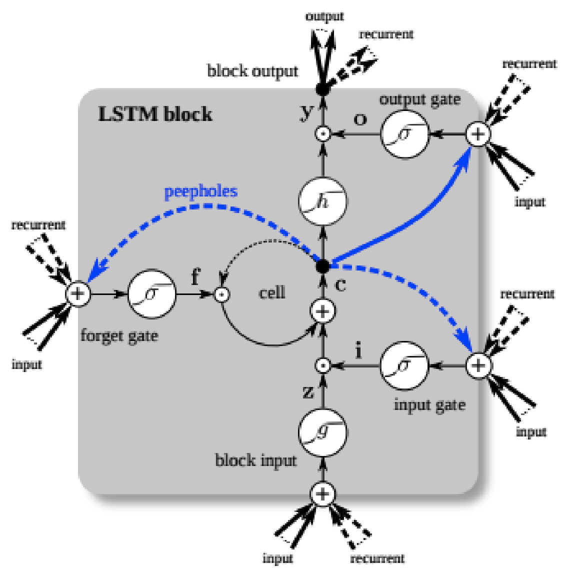

Figure 3.7: The process of judgment and calculation of LSTM

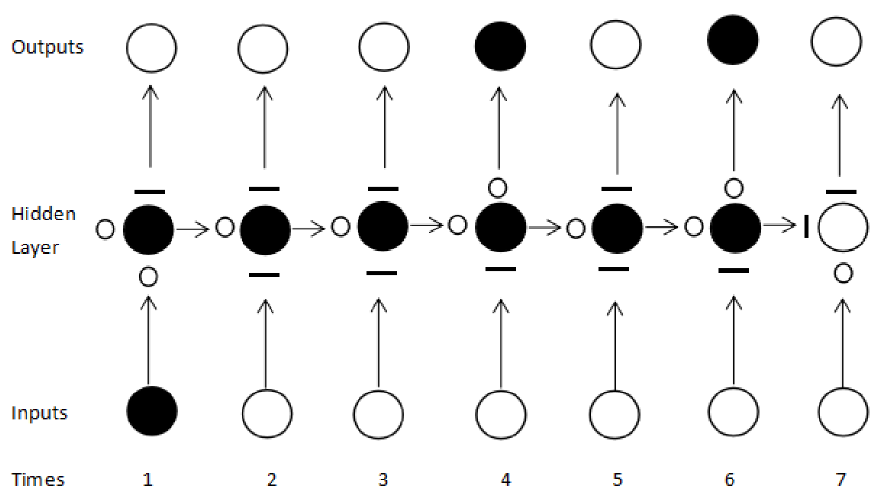

Figure 3.8: The operating principle of valve

state of the underlying cell. The mathematical expression is

$$f(t) = \sigma\left(W_f h^{(t-1)} + U_f x^{(t)} + b_f\right). \tag{3.39}$$

As can be seen from the figure, the input gate is involved of two parts. And the first part uses the sigmoid activation function, and the output is $i(t)$; and the second part uses the tanh activation function, and the output is $c'(t)$, and then we're going to multiply them and update the state of the cell. The mathematical expression is,

$$\begin{aligned} i^{(t)} &= \sigma\left(W_i h^{(t-1)} + U_i x^{(t)} + b_i\right), \\ c'^{(t)} &= \tanh\left(W_c h^{(t-1)} + U_c x^{(t)} + b_c\right). \end{aligned} \tag{3.40}$$

The mathematical expression for updating the cell state is,

$$\begin{aligned} i^{(t)} &= \sigma\left(W_i h^{(t-1)} + U_i x^{(t)} + b_i\right), \\ c'^{(t)} &= \tanh\left(W_c h^{(t-1)} + U_c x^{(t)} + b_c\right). \end{aligned} \tag{3.41}$$

Where $W_i, U_i, b_i, W_c, U_c, b_c$ are all the coefficients and bias of linear relationship, similar to that in RNN. An is a sigmoid activation function.

LSTM forward propagation algorithm Forward propagation algorithm of LSTM model. The LSTM model has two hidden states $h^{(t)}, C^{(t)}$ and the model parameters are four times as many as the RNN, there are $U_c, b_c, W_i, U_i, b_i, W_o, U_o, b_o, W_f, U_f, b_f, W_c$ more.

Forward propagation process:

Update the forgotten gate output,

$$f^{(t)} = \sigma\left(W_f h^{(t-1)} + U_f x^{(t)} + b_f\right). \tag{3.42}$$

Update the two part output of the input,

$$\begin{aligned} i^{(t)} &= \sigma\left(W_i h^{(t-1)} + U_i x^{(t)} + b_i\right), \\ c^{(t)} &= \tanh\left(W_c h^{(t-1)} + U_c x^{(t)} + b_c\right). \end{aligned} \tag{3.43}$$

Update cell status,

$$C^{(t)} = C^{(t-1)} \odot f^{(t)} + i^{(t)} \odot a^{(t)}. \tag{3.44}$$

Update the output of the output,

$$\begin{aligned} o^{(t)} &= \sigma\left(W_o h^{(t-1)} + U_o x^{(t)} + b_o\right), \\ h^{(t)} &= o^{(t)} \odot \tanh\left(C^{(t)}\right). \end{aligned} \tag{3.45}$$

Update the current time forecast output,

$$\hat{y}^{(t)} = \sigma\left(Vh^{(t)} + c\right). \tag{3.46}$$

LSTM back propagation algorithm Back propagation algorithm. The idea is consistent with the back propagation idea of RNN, and all parameters are iteratively updated by gradient descent method. The key is to compute the partial derivatives of all the parameters based on the function in loss.

In order to reduce the error of back propagation, we propagate step by step by hiding the gradient of the state $h(t)$ and $\delta^{(t)}$ There are two state functions $h^{(t)}$ and $C^{(t)}$, in LSTM and here we define two δ,

$$\begin{aligned} \delta(t)h &= \frac{\partial L}{\partial h^{(t)}}, \\ \delta(t)C &= \frac{\partial L}{\partial C^{(t)}}. \end{aligned} \tag{3.47}$$

For the sake of calculation, we divide the loss function $L(t)$ into two parts, the loss $l(t)$ at time t and the loss $L(t+1)$ after time t,

$$y = \begin{cases} l(t) + L(t+1) & (t < \tau), \\ l(t) & (t = \tau), \end{cases} \tag{3.48}$$

While the last moment $\tau, \delta_h^{(\tau)}$ and $\delta_C^{(\tau)}$ can be expressed as,

$$\begin{aligned} \delta_h^{(\tau)} &= \left(\frac{\partial O(\tau)}{\partial h(\tau)}\right)^T \frac{\partial L(\tau)}{\partial O(\tau)} = V^T\left(\hat{y}^{(\tau)} - y^{(\tau)}\right), \\ \delta_C^{(\tau)} &= \left(\frac{\partial h(\tau)}{\partial C(\tau)}\right)^T \frac{\partial L(\tau)}{\partial h(\tau)} = \delta_h^{(\tau)} \odot o^{(\tau)} \odot \left(1 - \tanh^2\left(C^{(\tau)}\right)\right), \end{aligned} \tag{3.49}$$

Then, we can get $\delta_h^{(t)}, \delta_C^{(t)}$ from $\delta_C^{(t+1)}, \delta_h^{(t+1)}$, and the gradient of $\delta_h^{(t)}$ is determined by the gradient error output at time t and the error after time t, it is,

$$\delta_h^{(t)} = \frac{\partial L}{\partial h^{(t)}} = \frac{\partial l^{(t)}}{\partial h(t)} + \left(\frac{\partial h^{(t+1)}}{\partial h^{(t)}}\right)^T \frac{\partial L^{(t+1)}}{\partial h^{(t-1)}} = V^T\left((y)^{(t)} - y^{(t)}\right) + \left(\frac{\partial h^{(t+1)}}{\partial h(t)}\right)^T \delta_h^{(t+1)}. \tag{3.50}$$

The difficulty of calculating the back propagation of the whole LSTM model is $\frac{\partial h^{(t-1)}}{\partial h(t)}$, and $h^{(t)} = o^{(t)} \odot \tanh\left(C^{(t)}\right)$. Where, $c^{(t)} = C^{(t-1)} \odot f^{(t)} + i^{(t)} \odot a^{(t)}$, and the results are as follows:

$$\Delta C = \rho^{(t+1)} \odot \left[1 - \tanh^2\left(C^{(t+1)}\right)\right], \tag{3.51}$$

$$\begin{aligned} \frac{\partial h^{(t-1)}}{\partial h(t)} &= W_o^T\left(o^{(t+1)} \odot \left(1 - o^{(t+1)}\right) \odot \tanh\left(C^{(t+1)}\right)\right) \\ &\quad + W_t^T\left[\Delta C \odot f^{(t+1)} \odot \left(1 - f^{(t+1)}\right) \odot C^{(t)}\right] \end{aligned}$$

$$+ W_c^T \Delta C \odot i^{(t+1)} \odot \left[1 - \left(c^{(t+1)}\right)^2\right]$$

$$+ W_t^T \left[\Delta C \odot c^{(t+1)} \odot i^{(t+1)} \odot \left(1 - i^{(t+1)}\right)\right]. \tag{3.52}$$

The reverse gradient error of $\delta_C^{(t)}$ is composed of the gradient error of the previous layer and the gradient error returned by the layer:

$$\begin{aligned} \delta_C^{(t)} &= \left(\frac{\partial C^{(t+1)}}{\partial C^{(t)}}\right)^T \frac{\partial L}{\partial C^{(t+1)}} + \left(\frac{\partial h^{(t)}}{\partial C^{(t)}}\right)^T \frac{\partial L}{\partial h^{(t)}} \\ &= \left(\frac{\partial C^{(t+1)}}{\partial C^{(t)}}\right)^T \delta_C^{(t+1)} + \delta_h^{(t)} \odot o^{(t)} \odot \left(1 - \tanh^2\left(C^{(t)}\right)\right) \\ &= \delta_C^{(t+1)} \odot f^{(t+1)} + \delta_h^{(t)} \odot o^{(t)} \odot \left(1 - \tanh^2\left(C^{(t)}\right)\right), \end{aligned} \tag{3.53}$$

And because of $\delta_h^{(t)}$ and $s_c^{(t)}$, it is easy to calculate all parameters if W_f is given, such as $U_f, b_f, W_c, U_c, b_c, W_i, U_i, b_i, W_o, U_o, b_o$.

$$\frac{\partial L}{\partial W_f} = \sum_{i=1}^{r} \left[\delta_c^{(t)} \odot C^{(t-1)} \odot f^{(t)} \odot \left(1 - f^{(t)}\right)\right] \left(h^{(t-1)}\right)^T. \tag{3.54}$$

3.2.2 Traditional prediction model

Many traditional models perform well in tourism prediction. We will focus on the two traditional prediction models to facilitate the comparison of our proposed models in this thesis. In the meantime, it is necessary to analyze the role of the scroll window.

3.2.2.1 Grey model

Grey model is a long-term prediction model, which processes random elements in the prediction system as gray data, and then finds out the internal law of the data. The required amount of data is small, and the prediction accuracy is high. The internal law of the system can be analyzed by a limited number of external elements representing the behavior characteristics of the system. And the grey model theory adopts the method of generating the behavior characteristic data of the system, processes the chaotic behavior characteristic data of the system, and discovers the internal law of the system from the chaotic phenomena, which is the uniqueness of the model. The gray model can be used to forecast the long term system behavior of periodic change and the system behavior of a non-periodic change because of its high adaptation. But the prediction effect for nonlinear data sample is poor.

Grey models utilize differential equations to characterize the complex system and to make prediction. For the grey systems, the GM(n, m) defines a grey model where m is the number of variables and n is the order of the difference equation.

The modeling process of GM$(1, n)$:

Suppose the system has characteristic data sequence:

$$X_1^{(0)} = \left(x_1^{(0)}(1), x_1^{(0)}(2), \cdots x_1^{(0)}(n)\right). \tag{3.55}$$

Sequence of related factors:

$$\begin{aligned} X_2^{(0)} &= \left(x_2^{(0)}(1), x_2^{(0)}(2), \cdots x_2^{(0)}(n)\right), \\ &\vdots \\ X_N^{(0)} &= \left(x_N^{(0)}(1), x_N^{(0)}(2), \cdots x_N^{(0)}(n)\right), \end{aligned} \tag{3.56}$$

Make $X_i^{(0)}(i = 1, 2, \cdots N)$ of The sequence of 1-AGO is $X_i^{(1)}$, where

$$X_i^{(1)}(k) = \sum_{k=1}^{n} x_i^{(0)}(k), \quad (i = 1, 2, \cdots n), \tag{3.57}$$

And then generating $X_i^{(1)}$ of the adjacent to the mean sequence $Z_1^{(1)}$, where

$$Z_1^{(1)}(k) = \frac{1}{2}\left[X_1^{(1)}(k) + X_1^{(1)}(k-1)\right], \quad k = 2, 3, \cdots n, \tag{3.58}$$

$x_1^{(0)}(k) + aZ_1^{(1)}(k) = \sum_{i=2}^{N} b_i x_1^{(1)}(k)$ is GM$(1, n)$ model. And for the GM$(1, n)$ model, a is the development coefficient, b_i is driving factor and $b_i x_i^{(1)}(k)$ is drive item, let

$$B = \begin{bmatrix} -Z^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) \\ -Z^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3) \\ \vdots & \vdots & \vdots & \vdots \\ -Z^{(1)}(n) & x_2^{(1)}(n) & \cdots & x_N^{(1)}(n) \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(n) \end{bmatrix}, \tag{3.59}$$

and let $\beta = (a, b_1, b_2, \cdots b_N)^T$, from the least square method, parameter estimation can be obtained: $X_i^{(1)}(i = 1, 2, \cdots N)$, when $X_i^{(1)}(i = 1, 2, \cdots N)$ has a little change, the approximate time corresponding formula is:

$$\hat{x}_1^{(1)}(k+1) = \left[x_1^{(0)}(1) - \frac{1}{a}\sum_{i=2}^{N} b_i x_i^{(1)}(k+1)\right] e^{-ak} + \frac{1}{a}\sum_{i=2}^{N} b_i x_i^{(1)}(k+1), \tag{3.60}$$

and the reduction formula is

$$\hat{x}_1(k+1) = \hat{x}_1(k+1) - \hat{x}_1(k), \tag{3.61}$$

The differential simulation equation is

$$\hat{x}_1^{(0)}(k) = -aZ_1^{(1)}(k) + \sum_{i=2}^{N} b_i \hat{x}_i(k). \tag{3.62}$$

3.2.2.2 Rolling window selection for gray model

The analysis of Rolling-window model for a time-series The stability of model over time and prediction accuracy of the model. A common model for time series assumes the coefficients are constant for time. And checking the instability is equivalent to checking if the coefficients are time invariant.

Select the size of scroll window m, i.e., that is, the number of consecutive observations per scroll window. And the size of the scroll window will rely on the sample size T, and the data periodicity. In general, for data collected at short intervals, a shorter scroll window size can be used, and for data collected at long intervals, a larger scroll window size can be used. A longer scroll window size produces a smoother scroll window estimate than a shorter scroll window size. Assuming that the increment between successive scrolling Windows is 1 period, and the entire data set is divided into $N = T – m + 1$ subsamples. The first scroll window contains observations from period 1 to m, the second scroll window contains observations from period 2 to $m + 1$, and so on. So there are some changes on the partition, for example, you can roll in four observations for quarterly data instead of one earlier. Let's estimate the model using each scroll window subsample. Plot the point to point confidence interval and each estimate value (i.e., $\hat{\theta} \pm 2[\hat{S}E(\hat{\theta})]$) to estimate how the exponent changes over time in the rolling window. Each parameter should fluctuate a little, but large trends or fluctuations indicate that the parameter may be time-varying.

Scroll window analysis predicts performance Suppose that there have data for all periods for the sample. We can use a rolling window for backtesting to check the predictive performance for the several time series models. The backtest is outlined by all steps.

Select the scroll window size m , that is, the number of consecutive observations per scroll window. The size of the scroll window depends on the sample size, T, and periodicity of the data. In general, for data collected at short intervals, a shorter scroll window size can be used, and for data collected at long intervals, a larger scroll window size can be used. A longer scroll window size produces a smoother scroll window estimate than a shorter scroll window size.

Choosing a prediction range, h. the prediction range depends on the application and periodicity of the data. The following shows how the scroll window divides the data set. If the increment in successive scrolling windows is 1 period, the $N = T – m + 1$ subsamples can get by dividing entire data set. Partitions are shown in the figure 3.9.

Estimate the each mode Estimate h-step-ahead in forecasts, and compute the prediction errors of each prediction, that is $e_{nj} = y_m - h + n + j - \hat{y}_{nj}$, where: e_{nj} is the forecast error of rolling window n for the j-step-ahead forecast. y is the response $\hat{y}_{nj}$ is the j-step-ahead prediction of rolling window subsample n.

Although the GM$(1, 1)$ is able to achieve good performance with small-sample observations, it could not capture the real-time trend because the whole-length time-series is used for training. A rolling window can be used [49–51] to address the problem and increase the forecasting accuracy. The idea of rolling window is to only use l latest samples for prediction where l is the window size. That means $\hat{x}^{(0)}(t+1)$ is predicted from $\left(x^{(0)}(n-l), x^{(0)}(n-l+1), \cdots, x^{(0)}(t)\right)$ instead of the whole training time series. The optimal window size l can be determined using the grid search algorithm which evaluates the performance of different window sizes on the one-step-ahead forecasting task [52–57].

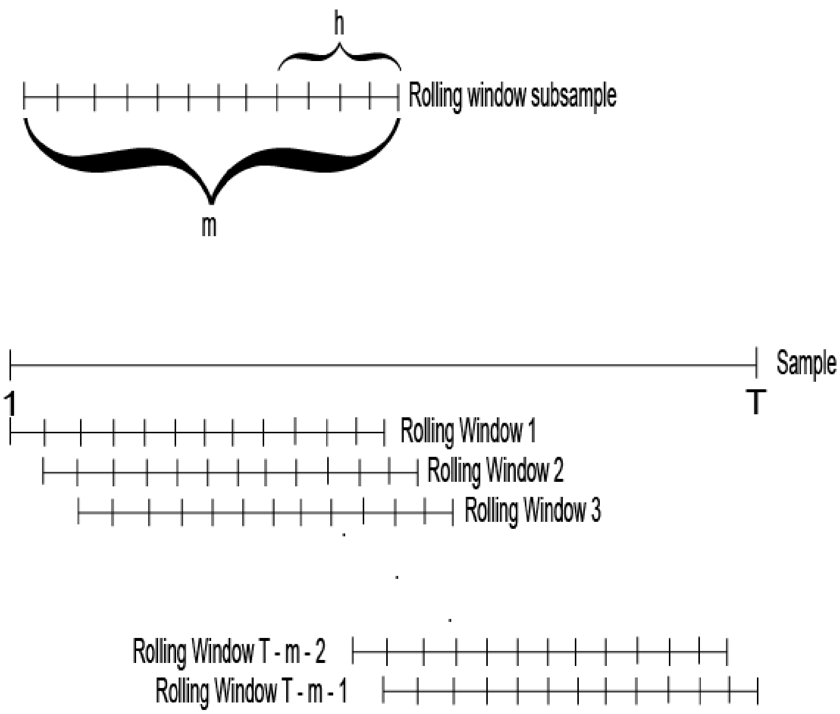

Figure 3.9: The architecture of scroll window

3.2.2.3 ARIMA model

The model of ARIMA was first introduced by Box and Jenkins in 1960s and is a classical time series model for tourist forecasting. ARIMA(Autoregressive Integrated Moving Average Model), also known as ARIMA(p, d, q), is a statistical prediction Model. In the process of prediction in economy, ARIMA model not only considers the dependence of economic phenomena on time series, but also considers the interference of random fluctuations based on autocorrelation analysis of time series. It is one of the widely used methods to predict the short-term trend of economic operation with high accuracy.

The basic program of the ARIMA model in forecast According to the graph of scatter, the autocorrelation function and the partial autocorrelation function graph of time series, the trend, the variance and the seasonal variation rule are tested by ADF unit root, and the stationarity of the sequence is identified. Generally speaking, the economic operation in time series is not a stationary series.

The nonstationary sequence is stabilized. If the data in sequence is in non-stationary and has a certain trend of growth or decline, the data will need to be processed differentially, if the data has heteroscedasticity, the data needs to be processed technically until the value of autocorrelation function and partial correlation function of the processed data are not significantly different from zero.

According to rules of the time series model in the recognition, the corresponding model will be established. If the autocorrelation function is tailed and the partial correlation function in the stationary sequence is truncated, we can draw the conclusion that the sequence is fit for the AR model. And if the partial correlation function of stationary sequence is tailed and the autocorrelation function is truncated, it can be concluded that the sequence is suitable for MA model. In a similar way, if the partial correlation function and autocorrelation function of stationary sequence are trailing, the sequence is suitable for ARMA model.

In the end, parameter estimation is carried out to test if it has statistical significance, carried out hypothesis test to diagnose if the residual sequence is white noise and using the tested model for prediction analysis.

The so-called ARIMA model means the model established by transforming the non-stationary time series into stationary time series and then regresses the dependent variable only to its present value and the lag value and lag value of the random error term. ARIMA model can be divided into four parts: moving average process (MA(q)), autoregressive process (AR(p)), autoregressive moving average process (ARMA(p, q)) and ARIMA(p, d, q).

We include this model to compare the performance with our hybrid AI model. A non-seasonal ARIMA can be written as:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t. \tag{3.63}$$

Where y'_t is the differenced series (it maybe have been differenced more than once), ϕ is AR coefficient; θ indicates MA coefficient. The 'predictors' on the right hand side include both lagged values of ARIMA(p, d, q) model, where p, d, q are the and lagged errors. We call this an auto-regressive term, the integrated term and the moving-average term, respectively. The auto-regressive term illustrates the lags of time series and the moving-average terms refers to the lags of forecast errors. The integrated term refers to the difference levels to obtain a stationary time series. The values of p, d, q are determined by the statistical properties of the time series.

Once it starts combining components to form more complicated models in this way, it is quite easier to work together the backshift notation. For example, equation (3.63) can be written in backshift notation as equation (3.64).

$$\begin{array}{ccccc} (1-\phi_1 B - \cdots - \phi_p B^p) & (1-B)^d y_t & = & c + (1+\theta_1 B + \cdots + \theta_q B^q)\,. \\ \uparrow & \uparrow & & \uparrow \\ \text{AR}(p) & d \text{ differences} & & \text{MA}(q) \end{array} \tag{3.64}$$

R uses a slightly different parameterisation:

$$(1-\phi_1 B - \cdots - \phi_p B^p)\,(y'_t - \mu) = (1+\theta_1 B + \cdots + \theta_q B^q)\,\varepsilon_t. \tag{3.65}$$

where $y'_t = (1-B)^d y_t$ and μ is the mean of y'_t.To convert to the form is given by (3.65). Set $c = \mu\,(1-\phi_1 - \cdots - \phi_p)$. [57–63]

Some common models

1. **ARIMA(0,1,0): the random walk.** When $d = 1$ and p and q are 0, it is called random walk, as shown in the figure. The position of each moment is only related to the position of

the previous moment. The prediction formula is as follows:

$$\widehat{Y}_t = \mu + Y_{t-1}. \tag{3.66}$$

2. **ARIMA(1,0,0): the first-order autoregressive model** When $p = 1$, $d = 0$, $q = 0$. That the temporal data is stable and autocorrelated. The Y value at one time is only related to the Y value at the previous time. The prediction formula is as follows:

$$\hat{Y}_t = \mu + \phi_1 * Y_{t-1}. \tag{3.67}$$

where, $\phi \in [-1, 1]$ is a slope coefficient.

3. **ARIMA(1,1,0): the differenced first-order autoregressive model.** When $p = 1$, $d = 1$, $q = 0$. It indicates that the time series data are stable and autoregressive after first-order differential differentiation. That is, the difference at one time (y) is only related to the difference at the previous time. The prediction formula is as follows:

$$\hat{y}_t = \mu + \alpha_1 * e_{t-1}, \tag{3.68}$$

Combined with the definition of first-order difference, it can also be expressed as:

$$\hat{Y}_t - Y_{t-1} = \mu + \phi_1 * (Y_{t-1} - Y_{t-2}) \quad \text{or} \quad \hat{Y}_t = \mu + Y_{t-1} + \phi_1 * (Y_{t-1} - Y_{t-2}). \tag{3.69}$$

4. **ARIMA(0,1,1): the simple exponential smoothing with growth.** When $p = 0$, $d = 1$, $q = 1$. It shows that the data are stable and moving average after first-order difference. That is, the difference of the estimated value at one time is related to the prediction error at the previous time.

$$\hat{y}_t = \mu + \alpha_1 * e_{t-1},$$

where,

$$\hat{y}_t = \hat{Y}_t - \hat{Y}_{t-1}, e_{t-1} = Y_{t-1} - \hat{Y}_{t-1},$$

set $\theta_1 = 1 - \alpha_1$

$$\widehat{Y}_t = \mu + \widehat{Y}_{t-1} + \alpha_1 \left(Y_{t-1} - \hat{Y}_{t-1} \right). \tag{3.70}$$

5. **ARIMA(2,1,2) can be expressed as follow:**

$$\hat{y}_t = \mu + \phi_1 * y_{t-1} + \phi_2 * y_{t-2} - \theta_1 * e_{t-1} - \theta_2 * e_{t-2}$$

or

$$\begin{aligned}\hat{Y}_t = \mu &+ \phi_1 * (Y_{t-1} - Y_{t-2}) + \phi_2 * (Y_{t-2} - Y_{t-3}) \\ &- \theta_1 * \left(Y_{t-1} - \hat{Y}_{t-1} \right) - \theta_2 * \left(Y_{t-2} - \hat{Y}_{t-2} \right),\end{aligned} \tag{3.71}$$

and ARIMA$(2,2,2)$ can be expressed as

$$\widehat{Y}_t = \mu + \widehat{Y}_{t-1} + \alpha_1 \left(Y_{t-1} - \hat{Y}_{t-1}\right)$$

or

$$\begin{aligned}\hat{Y}_t = \mu &+ \phi_1 * (Y_{t-1} - 2Y_{t-2} + Y_{t-3}) + \phi_2 * (Y_{t-2} - 2Y_{t-3} + Y_{t-4}) \\ &- \theta_1 * \left(Y_{t-1} - \hat{Y}_{t-1}\right) - \theta_2 * \left(Y_{t-2} - \hat{Y}_{t-2}\right).\end{aligned} \tag{3.72}$$

The results of ARIMA will be analyzed in the next chapter for the research on the problems of tourism.

3.2.3 GM-LSTM model with rolling windows

To better predict the arrivals in number of in Xi 'an, this chapter will study the relevant models and algorithms and put forward an appropriate model: GM-LSTM model.

3.2.3.1 Method framework

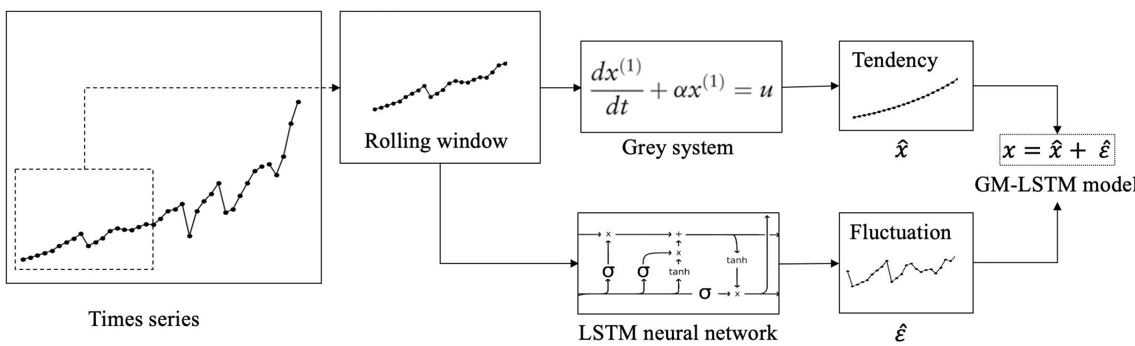

Figure 3.10: Overview framework of hybrid GM-LSTM model

It shows in Fig. 3.10 that the general trend of the data is predicted by the rolling window GM model, and LSTM neural network model is used to decompose and fit the fluctuations, and then their results of the rolling window GM model, and LSTM of the neural network model are finally added up. The specific process is as follows.

The annual tourism demand time series usually show an overall trend with fluctuation, which can be seen as observations from complex systems. For instance, thedemand of annual tourism is often affected by various factors including tourism price, exchange rate and security. To capture the real-time system state and make adaptive prediction, we use a rolling window to select a segment time series for analysis and prediction. The rolling window is shown on the left of Figure 3.10. The first step is to capture the window data, and then move it backward once the relevant calculation is done to get the window data. The size of the window is described later. Then, the GM model is used to predict the general trend, and the tendency $\hat{x}$ is first extracted using first-order gray model from the rolling windows. Then the residual fluctuation $\hat{\varepsilon}$ is decomposed from the time series and fitted using LSTM neural network. In the end, by combining the results of GM model and LSTM neural network model, the result x can be predicted effectively. After learning from the history data, one-step-ahead prediction can be made with a rolling window. The framework is demonstrated in Figure 3.10. [64–78]. The related models and algorithms of this hybrid framework is introduced in following sections.

3.2.3.2 GM-LSTM model

As discussed above, LSTM was first proposed by Hochreiter and Schmidhuber in 1997 to solve the convergence problems faced by traditional neural networks in time sequence prediction. The general structure of the LSTM neural network is demonstrated in Figure 3.11(A).

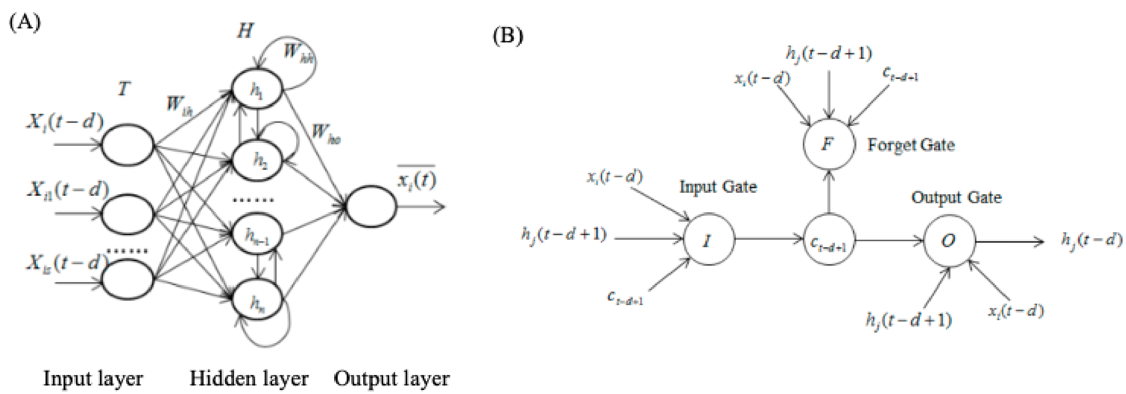

Figure 3.11: LSTM model of time series prediction (A) and structure of LSTM neuron (B)

The model is made up three layers: the output layer, the hidden layer and the input layer. The inputs are the observations of x at previous $t-l$ historical moments $T = (x(t-l), x(t-l+1), \cdots, x(t))$ and the ouput is the forecasting value of x at $(t+1)$-th moment denoted as $\hat{x}(t+1)$. The neurons of the hidden layer are denoted as $H = (h_1, h_2, \cdots, h_j, \cdots, h_n)$ where h_j is the j -th neuron of the hidden layer. The weight within the hidden layer is denoted as W_{hh}. The weights between the input layer and the hidden layer is denoted as W_{ih} and the weights between the output layer and hidden layer is W_{ho}. The model is calculated as:

$$h_n = H\left(W_{ih}T + W_{hh}h_{n-1} + b_h\right), \tag{3.73}$$

$$\hat{x}(t+1) = W_{ho}H + b_y. \tag{3.74}$$

where b_h is the bias vector in the hidden layer and b_y is the bias vector in the output layer.

Each neuron in Figure 3.11(A) is composed of three gates: the forgetting gate, the output gate and the input gate as shown in Figure 3.11(B). The gate of input determines that information to be the input of the current moment state c_{t-d} according to $x(t-d)$, c_{t-d+1} and $h_j(t-d+1)$. The gate of the output determines the output of the current moment state according to the $c_{t-d,}h_j(t-l+1)$ and $x(t-d)$. The forget gate determines which information to be forgotten according to the current input $x(t-d)$ the last moment state of the neuron c_{t-d+1} and the last moment output the of j-th neuron $h_j(t-d+1)$. The calculation of the input gate and the forget gate are:

$$I = \sigma\left(x_i(t-d) + W_{Ic}c_{t-d+1} + W_{Ih}h_j(t-d+1) + b_I\right), \tag{3.75}$$

$$F = \sigma\left(W_{fi}x_i(t-d) + W_{fc}c_{t-d+1} + W_{fh}h_j(t-d+1) + b_f\right). \tag{3.76}$$

The update state of the neuron is calculated as:

$$c_t = F * c_{t-d+1} + I * g\left(W_{ci}x_i(t-d) + W_{ch}h_j(t-d+1) + W_{cc}c_{t-d+1} + b_c\right), \tag{3.77}$$

The calculation of the output gate is:

$$O = \sigma\left(W_{oi}x_i(t-d) + W_{hh}h_j(t-d+1) + W_{oc}c_{t-d+1} + b_o\right). \tag{3.78}$$

The selective memory of the function in LSTM network model implemented by the gating mechanism makes LSTM more suitable to deal with time sequence forecast than the traditional neural networks [79–85]. Integrating the GM and LSTM models we get the GM-LSTM model.

We integrate the first-order gray model and LSTM neural network with a rolling mechanism. Let $(x(1), x(2), \cdots, x(t))$ be the full time series data for training. As described in Section 3.2, we first extract training samples $\{S_k\}$ for one-step prediction:

$$S_k = [x(k), x(k+1), \cdots, x(k+l)], \quad k = 1, 2, \cdots, \tag{3.79}$$

where S_k is the k-th training sample including a l-length time series data. The corresponding one-step-ahead data $x(k+l+1)$ is the regression target. A first-order gray model is used to fit the sample and make a prediction $\hat{x}(k+l+1)$ The residual error of gray model prediction for the k-th training sample is calculated as:

$$\varepsilon_k = x(k+l) - \hat{x}(k+l+1), \quad k = 1, 2, \cdots, t-l. \tag{3.80}$$

Therefore, we can obtain the training label ε_k of the sample S_k for LSTM training.

In the utilization stage, to predict the data $x(t+1)$, we get the latest sample series $S_{t-l} = [x(t-l), x(t-l+1), \cdots, x(t)]$. Then a trend prediction $\hat{x}(t+1)$ can be made using gray model. The residual fluctuation $\hat{\varepsilon}_{t-l}$ is also predicted using LSTM given the latest sample series. Finally, the combination of the two AI models gives the one-step-ahead prediction:

$$\tilde{x}(t+1) = \hat{x}(t+1) + \hat{\varepsilon}_{t-l}. \tag{3.81}$$

After one-step-ahead prediction is obtained, the prediction of subsequent timesteps can be made iteratively using the same procedure [86–90].

GM model can use less data to predict the trend of time series. Theoretically, four data can be used to predict the trend. LSTM, on the other hand, is well suitable for dealing with problems that are highly related to the time series, and its next time step will have a weight connected in parallel to itself to copy the true value for its own state and the accumulated the external signal. However this kind of self-connection is controlled by a multiplicative gate that another unit learns and decides when to clear the memory. So it's good for predicting the residual between the trend and the expected value.

For the tourism data of countries along the route of One Belt and One Road to Xi 'an, there has a less in number, big fluctuation and high non-linear characteristics. Combining GM model with LSTM can better solve these problems: GM model predicts the overall trend of change, LSTM model captures and processes the fluctuation information, and the rolling window processing method is able to make full use of the existing data to predict a more reliable result.

For the well-comparison, the results are divided into three parts: the results of the related neural network model, the results of the traditional model and the results of the GM-LSTM model. We will discuss the results in next chapter.

3.3 Construction of credibility on tourism model: the case of data mining in tourism

A lot of comments now is called the merchant navy to campaign to attract consumers, or that is to deceive consumers, the real sound part to tourists issued false information is submerged in the boundless sea of these comments. And some of the real scenic spots are also deliberately blackened by other competitors, which eventually make no one really worth traveling. More and more people like to give a first impression to tourist spots through online reviews, but excessive information and widely different comments make consumers feel more confused. How to judge the credibility of information in these comments is the establishment of this model of the mind, the data mining algorithm of tourism online review credibility model can greatly improve the information filtering technology based on operation of real information extraction is greatly improved, we use it as a breakthrough point to establish the model:

From here you can see that, when consumers provide initial data by data mining algorithms for the calculation, and then given a set of data, through a series of processing the range of target data reduction, data mining on the selected data by other test methods, hidden random information, finally this evaluation, customers are satisfied with the results. Based on big data, JSP, JDBC and so on, web works as a way. Its work is divided into three levels, namely, data level, logic level and expression level. Data analysis is used to find relevant points, and finally the results are expressed. Intelligent information processing module includes analysis, data mining, intelligent screening, reliability evaluation. When consumers in the engine are consumer information system to intelligent credibility evaluation calculation of online review scores, through the comment number and related consumer information inspection to determine the authenticity of the comments and color commentators, eventually hit the credibility of the score for comments continue this cycle of operation, and ultimately the formation of a set of data, according to consumers are given the intention to select similar comments feedback to consumers, to achieve purpose. The module of the system in each division operation, and data mining algorithms interspersed among them, form a

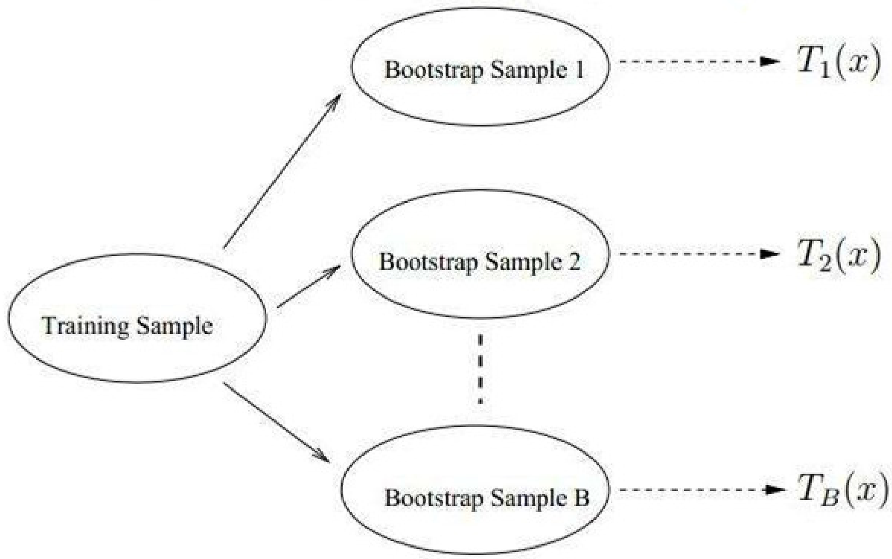

Figure 3.12: Data mining process

complementary, interactive result evaluation model.

$$\frac{\partial \varepsilon}{\partial {W_i}^{(j)}} \overset{def}{\Rightarrow} \left[\frac{\partial \varepsilon}{\partial {w_{1i}}^{(j)}}, ..., \frac{\partial \varepsilon}{\partial {w_{1i}}^{(j)}} \frac{\partial \varepsilon}{\partial {w_{m_{j-1}+1}}^{(j)}} \right], \tag{3.82}$$

$$Z_c = \frac{|T - n(n+1)/4|}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum\left(t_j^3 - t_j\right)}{48}}}. \tag{3.83}$$

First, the weight matrix of the index is obtained by the analytic hierarchy process (AHP). Among them, T is the weight value of each index and according to the consistency check of (2) formula, the credibility is analyzed. Then, the function Z in the weighted average type comprehensive evaluation model is established to represent the c individual trustworthiness of the n generation. The t represents the data mining error and the T represents the weight of the system data processing structure. The j between (0, 1), is an adjustment parameter that is responsible for controlling the proportion of data mining error and structural complexity in the Z function. In fact, some nodes have no effect on the reliability investigation and calculation, and the influence on the final result is also low. The data mining set can be optimized by the next type. That formula is second level of the individual in the position on the K value.

$$\varepsilon_j(k) = |x_0(k) - x_i(k)| + 0.5 \max_i |x_0(k) - x_i(k)|, \tag{3.84}$$

$$r_i = \frac{1}{N} \sum_{k=1}^{n} \varepsilon_i(k). \tag{3.85}$$

By testing the calculated data of the upper form, if we meet the consumer's credibility and accept the scope, we will export the data, and then test the credibility of the data to evaluate the data value and save the data. If it is not satisfied, the algorithm model calculation step is returned to continue to be calculated and analyzed, or there is no information of high credibility in the data. The algorithm we designed is very suitable for the application of online commentary credibility selection, since it ensures the speed of calculation and the accuracy of the calculated data and it is also a more successful algorithm for combining applications. In this way, we can raise the speed of our data search, as shown in the following table, for the speed and classification of data search:

Because the main step of the finite element analysis is to divide the problem, which is the core of our calculation. For example, the calculation of the following formula:

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 \cdots\cdots, \tag{3.86}$$

This calculation method that transforms the trigonometric function into the one element n equation is a computational method of regional discretization. Then we need to determine the state variable and the control method. This part is the step of solving the differential equation. That is to solve the differential equation by using the reminding of the finite element analysis. Generally, in order to make computer programming convenient, our formula is improved to the following form:

$$\begin{cases} T_1 = y_i + hf\left(x_i, y_i\right) \\ T_2 = y_i + hf\left(x_{i+1}, y_i\right). \end{cases} \tag{3.87}$$

Reliability Evaluation Model of Online Tourist Reviews Based on Data Mining

In order to reflect the whole system operation program, the above image is a data mining algorithm flow chart based on Web. Through a large amount of information, we can combine progressive functions with suitable functions. By weighting variables, we select the nearest average value of nodes in a class as a data point, and the specific way is: given an initial data set, the clustering number and the initial cluster center, and then calculated each data to the initial cluster center distance, and the distance of each data node closest to them from the nearest node data set is formed, and then calculated the data set to the initial cluster center distance to seek closer point; Through the consumer tourism point of intention given by the consumer, web is selected as the platform, and the comments are regarded as the initial data; the data can be seen as a matrix that surrounds the central cluster, which is converted into a difference degree matrix by clustering analysis; while the consumer intention is the initial cluster center, and these evaluations which have higher credibility will be close to the center and through the combined data set and continuous screening, these high credibility evaluations will be chose. This is based on the operation process of data mining tourism online review credibility model algorithm. After we carry out the above calculation, the whole data mining system theory of the tourism online review credibility evaluation model algorithm part even if the calculation is completed. Then we will be in accordance with the relevant knowledge of key factors are divided, as shown in the following table:

The process of making up a lot in the model construction is the construction of the weighted average comprehensive evaluation model. In this part of the construction, we use the related algorithms to build the model. This part is much more calculated by using resources and formulas, so that the time takes a lot of time. In a lot of basic data, we use formulas to calculate the data one by one too much, so we adopt the technology of programming. In formulae defined to a particular program, we use the powerful computing power of the computer to analyze data for us. The second is the construction of the weight matrix and the single factor matrix. In these two aspects, we use matrix equation to calculate the basic data processed in the previous step, and divide the weight value of the factors that affect the degree. We evaluate each factor from the mathematical point of view, and try to ensure the scientific nature of the whole result.

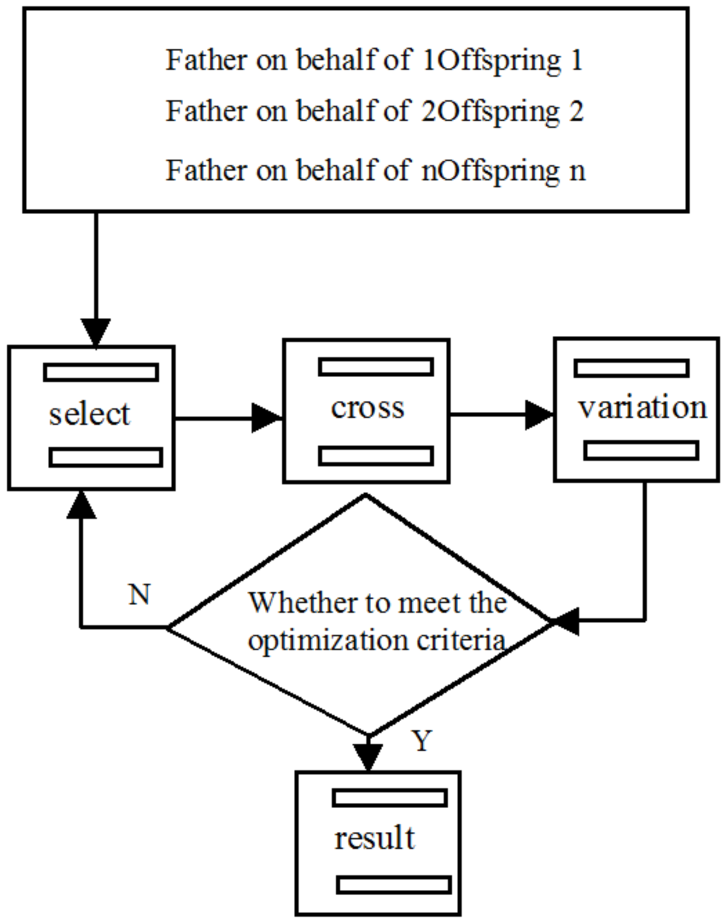

Figure 3.13: Reliability analysis of tourism online evaluation

In addition, in the evaluation of evaluation standards, we adopted the opinions of relevant experts, thus ensuring the scientificity of the evaluation standards. The calculation of the reliability model of travel online evaluation based on data mining shows that the expectations of consumers are very high at the time of selection. They want to explore the true evaluation of information through this system to determine where to go, and the data mining algorithm really does this. Consumer satisfaction has reached a high standard, and few have not expressed opinions. Most people think that the model really simulates their judgment index, which is a feasible model.

Chapter 4

Experiments

4.1 Dataset description

Forecasting data. The data of inbound tourists to Xi'an, including information related to inbound tourist arrivals and foreign currency income, has being full collected and researched. The total numbers of tourist arrivals Nearly 30 years and that of top 10 countries is collected. At the same time, the total numbers of tourists interested in top 10 spots, such as one of them, the 'Terra Cotta Warriors' museum in Xi'an are collected; All statistics of income and consumption coming from the tourists in transportation (plane, train and bus), accommodation, catering, entertainment, and consumption, such as shopping and communication of data, are carried out. Take hotel for an example, the top 10 star-hotel check-in and consumption situation of the data collection in this research has been counted. The above work has provided a good data support for this study, in the meantime, also lay a well foundation for the later research. In addition, for the research of prediction problem, it is necessary to introduce some more effective prediction model. In the following content, we are going to propose a new model together with these models to predict the same problems so as to compare with the new model' results in this study .

As described above, Xi'an is screened as the location of experimental research to evaluate the effectiveness of the presented forecasting scheme. Xi'an is a well-known tourist city in China and attracting a large number of tourists from the whole world every year. The data of Xi'an international tourist arrivals from 1980 to 2018 are selected to assess the effectiveness of our hybrid AI forecasting model. The number of international tourist arrivals for each year is collected from the government of Xi'an and listed in Table 4.1.

For convenience of comparison, the models of the neural network, the traditional models, and the model combined with the neural network model and the traditional model proposed in this thesis all of them use the same dataset, including training dataset and prediction dataset (also called held-out dataset).

4.2 Evaluation

The data from 1980-2013 was applied as training set and the data from 2014-2018 are the held-out test set to assess the performance of different models. In this study, three criteria are used for

Table 4.1: International tourist arrivals in Xi'an from 1980-2018 (Unit: ten thousand)

Year	Arrivals	Year	Arrivals
1980	4.00	2000	65.03
1981	6.71	2001	67.20
1982	9.09	2002	74.13
1983	12.38	2003	33.66
1984	15.13	2004	65.03
1985	21.15	2005	77.56
1986	25.78	2006	86.73
1987	30.15	2007	100.01
1988	36.58	2008	63.20
1989	21.20	2009	67.29
1990	25.88	2010	84.18
1991	31.00	2011	100.23
1992	40.16	2012	115.35
1993	43.50	2013	121.11
1994	41.49	Test data	
1995	41.35	2014	124.23
1996	45.39	2015	110.72
1997	48.53	2016	133.88
1998	47.98	2017	175.13
1999	55.41	2018	202.75

evaluating the performance: the mean square error (MSE), the absolute mean error (MAE), and the mean absolute percentage error (MAPE) [91–95] which are defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2, \tag{4.1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i|, \tag{4.2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - \hat{x}_i}{x_i}\right| \times 100\%. \tag{4.3}$$

4.3 Model implementation

All models are implemented using the same simulation software, using the same data, the same training set and the same prediction set to compare the effectiveness of the models.

4.3.1 The models related to neural network implementation

All the models were implemented and evaluated in open-source Python 3.7. The implementation details are described as follows:

4.3.1.1 SDNM implementation

When accomplishing the reconstruction in phase space and the index of maximum Lyapunov, the prediction tourist numbers can be carried out the according to SDNM described above. First, we all the time series data are divided into two parts: one of them was for the set of training data , and the other part was used to verify the accuracy of the prediction as a text set. And the weight W_{ij} was optimized by BP-like learning method. The synaptic threshold θ_{ij} SDNM until the condition of the learning termination was met. In this research, the Lmax of maximum learning epoch was used as the condition for termination. Finally, we use the output prediction results in some assessment methods.

Calculation time delay tau take the first minimum of mutual information value. The mutual information $I(y, y_\tau)$ between the two time series $y = (y_{t1}, y_{t2}, \cdots, y_{tN})$ and $y_\tau = (y_{t1+\tau}, y_{t2+\tau}, \cdots, y_{tN+\tau})$ is the bits of average value, where y is predicted by measurement y_τ, ang $I(y, y_\tau)$ can be expressed as,

$$I(\tau) = I(y, y_\tau) = H(y) + H(y_\tau) - H(y, y_\tau), \tag{4.4}$$

Where, $H(y)$ and $H(y_\tau)$ are the entropy of yand y_τ, respectively and the mutual entropy between y and y_τ is $H(y, y_\tau)$.

$H(y, y_\tau)$ is the mutual entropy between y and y_τ.

Input time series data.

Make x_t was the one-dimensional time series of tourism at time t, $(t = 1, 2, \cdots)$. First , x_t was input and processed using the method of normalization to the range of $[0, 1]$ according to Eq.(4.4).

$$y_t = \frac{x_t - MIN(x_t)}{MAX(x_t) - MIN(x_t)}. \tag{4.5}$$

Where, y_t was the standardized data of alleviating the problem for inconsistent data measures in different time series, and MAX (MIN) returns the maximum value (minimum value) of the vector.

An appropriate dimension m in embedding and time delay τ are used for transforming phase space. The algorithm proposed by Grassberger Procaccia is used to determine the dimension m in embedding and interactive information function was used to calculate the time delay τ. For the normalized time series y_t, $(t = 1, \cdots, N)$, and the phase space reconstructed by us could be expressed as a matrix $x(P, T)$.

$$P = \begin{pmatrix} y_1 & y_2 & \cdots & y_{N-1-\tau(m-1)} \\ y_{1+\tau} & y_{2+\tau} & \cdots & y_{N-1-\tau(m-2)} \\ & & \cdots & \\ y_{1+\tau(m-1)} & y_{2+\tau(m-1)} & \cdots & y_{N-1} \end{pmatrix}, \tag{4.6}$$

$$T = \left(y_{2+\tau(m-1)}, y_{3+\tau(m-1)}, \ldots, y_N\right). \tag{4.7}$$

4.3.1.2 RBFNN implementation

Data preprocessing and experimental setting to improve the forecasting accuracy, all used predicting data are normalized and scaled in the range of $[0, 1]$ by applying to the following Eq.(4.7).

$$y_k = \frac{y_{ok} - \min(\boldsymbol{y}_{ok})}{\max(\boldsymbol{y}_{ok}) - \min(\boldsymbol{y}_{ok})}, \tag{4.8}$$

where y_{ok} is the real value before scaling, and y_k is the value after scaled down. After the forecasting process is completed, the forecasting performance is tested by scaling down, and all the outputs for their actual values by using the following Eq.(4.9).

$$\widehat{y}_{ok} = \min\left(\boldsymbol{y}_{ok}\right) + \widehat{y}_k \times \left[\max\left(\boldsymbol{y}_{ok}\right) - \min\left(\boldsymbol{y}_{ok}\right)\right], \tag{4.9}$$

where $\widehat{y}_{ok}$ is the forecasting actual value and $\widehat{y}_k$ is the forecasting scaled value.

The nonlinear filters including the IEKF and the IUKF are based on the mean squared error (MSE) principle, so their average MSE (AMSE) are used for model comparison of evaluation.

The value of AMSE can be defined as:

$$\text{AMSE} = \frac{\sum_{k=1}^{M}\left(y_o k - \widehat{y}_o k\right)^2}{M}. \tag{4.10}$$

For the computing complexity problem, the open-source Python 3.7. in our work are performed, and their comparison results are provided as well.

A RBFNN, which constitutes ten inputs, ten centers and one output, are trained with the IEKF and the IUKF in an online fashion. Each element of the central vector and the weight matrix is initialized with 0.1, and their IEKF's initialized value of covariance matrices are $P_0 = 10I$, $Q_0 = 0.01I$ and $R_0 = 0.01I$ (where I is the corresponding dimensions identity matrix), respectively. The forecasting results of the average are across a Monte Carlo simulation run from 50 to a fair prediction comparison.

4.3.1.3 LSTM model implementation

RNN can think of it as a deep feedforward neural network where the same weight are shared by all layers. In spite of their main purpose is going to learn about long-term dependencies, both the empirical and theoretical evidence suggests it is hard to learn and retain information over time.

LSTM is a fair goog variant model of RNN, which inherits the most RNN models in characteristics and solves the gradient disappearance problem caused by the gradual decrease of gradient in the process of gradient back propagation. It really simulates or represents the cognitive processes of human behavior, logical development, and neural tissue. LSTM has become a very popular research model in RNN and even in the deep learning framework, which has received extensive research and attention. When it comes to the tasks of language-processing, LSTM is an ideal tool for dealing with the highly time-series-dependent problems. LSTM, a special unit, which is called memory cells, similar to the accumulator and gating neurons: it the next time step will have a weight in parallel to itself, copy their own state of real value and the accumulation of external signal, but that since the connection is made by another unit of learning and to decide when to remove the memory contents [96–99].

To solve our problem, an idea arose to increase network storage. LSTM (long short-termmemory networks) with a special implicit unit was first proposed. A special unit with memory cells similar to the accumulator and gating neurons: it the next time step will have a weight in parallel to itself, copy their own state of real value and the accumulation of external signal, but that since the connection is made by another unit of learning and to decide when to remove the memory contents by practice.

The LSTM network whereafter proved to be the more efficient than traditional RNNs model, especially when there were several layers in each time step, and the whole speech recognition system was able to transcribed the acoustics into character sequences in a completely consistent manner. Current LSTM networks and related gated units are used to encode and decode networks also, and perform well in machine translation.

4.3.2 Traditional prediction model implementation

4.3.2.1 Grey model

Grey system has relativity and universality. The grayscale of the system is different for different objects. As a practical matter, gray systems are abundant in the world. Grey forecast model is a method to forecast the system with many uncertain factors. By identifying the degree of difference between the development trend of system factors, it generates and processes the original data, and conducts correlation analysis to find the law of system change, and the data sequence with strong regularity is generated, and then the corresponding differential equation model is established, so as to forecast the future development trend of events [100–105]. In general, a grey prediction model is constructed by using a series of quantitative values observed in the same time distance to predict the quantity of the peculiarity at a certain time in the future, or to reach a certain quantity characteristic at the certain time.

The GM model has the following advantages: it does not need a great number of sample; and samples do not need a regular distribution; the calculation work is small; and the quantitative results of analysis will not be inconsistent with the qualitative results in analysis; it can be used for long term, medium and short-term prediction; and the grey prediction accuracy is high [106–108].

First, the classical GM$(1,1)$ gray model was implemented to fit the whole training set and made predictions on the test set. Then a rolling grey model with a rolling window is implemented on the training set. As described in Section 3.2.2.2, the best rolling window length was determined by a grid search within the range of $[4, 15]$. The optimal window length was decided by the MAPE of one-step-ahead prediction on the set of training . Then the prediction on the set of test was made through sliding the rolling window.

The GM$(1,1)$ is one of the most popular grey model using first-order differential equation for uni-variate forecasting. Let the raw time series be

$$x^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(j), \cdots, x^{(0)}(t)\right). \tag{4.11}$$

where $x^{(0)}(j)$ is the j-th observation and n is the total number of training data. The gray model is to predict the value $x^{(0)}(t+1)$. The basic procedure of GM$(1,1)$ is summarized as follows:

Step 1: Create one-order Accumulated Generating Operation (1-AGO) sequence:

$$x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(j), \cdots, x^{(1)}(t)\right). \tag{4.12}$$

where $x^{(1)}(k) = \sum_{j=1}^{k} x^{(0)}(j)$, for $k = 1, 2, \cdots, t$. The purpose of this step is to obtain a time series more regular than the original data, which helps to capture the tendency.

Step 2: Build the first-order ordinary differential equation of $x^{(1)}$:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u. \tag{4.13}$$

where a and u are called the developing coefficient and the gray input, respectively.

Step 3: Estimate the parameters a and u using least square method:

$$\begin{bmatrix} a \\ u \end{bmatrix} = \left(\boldsymbol{B}^T \boldsymbol{B}\right)^{-1} \boldsymbol{Y}_N, \tag{4.14}$$

where

$$B = \begin{bmatrix} -0.5\left(x^{(1)}(1) + x^{(1)}(2)\right) & 1 \\ -0.5\left(x^{(1)}(2) + x^{(1)}(3)\right) & 1 \\ \cdots & \cdots \\ -0.5\left(x^{(1)}(n-1) + x^{(1)}(n)\right) & 1 \end{bmatrix}, \tag{4.15}$$

and

$$\boldsymbol{Y}_N = \left[x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)\right]^T. \tag{4.16}$$

Step 4: Forecast model.

$$\hat{x}^{(1)}(t+1) = \left(x^{(0)}(1) - \frac{u}{a}\right) e^{-an} + \frac{u}{a}, \tag{4.17}$$

And

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k). \tag{4.18}$$

where $\hat{x}^{(0)}(k)$ is the prediction at k-th time [109].

4.3.2.2 Rolling model

The basic principle of the rolling window protocol is that at any time, the sender maintains a serial number of frames allowed to be sent, which is called the sending window. At the same time, the receiver also maintains a continuous number of allowable frames, called the receive window. The upper and lower bounds of the sending and receiving window Numbers do not have to be the same, and even the sizes can be different. Different sliding window protocol window size is generally different. The sequence number in the sender window represents those frames that have been sent but have not yet been confirmed, or those that can be sent.

In the above program, i points to the head of the array, j points to the end of the array, and k serves as the index of the characters in the substring. At the beginning of each comparison, k is initialized with the value of i, and when the duplicate character is found, the value of $k + 1$ is assigned to i, that is, the left side of the window is moved directly to the next character position of the duplicate character. A sliding window on the right side of each right, neutron string contains the window if the window on the right side of the next character, on the left side of the sliding one or multiple compared with enumeration method, the use of repeating characters in a substring location, directly to the window on the left side of the jump to the next position of the characters, each check out repeated reduced $k-i$ substring self-inspection. Similar to the longest unrepeated substring above, if the enumeration method is used, assuming the length of the array is L, then it is required to find $L-n+1$ and, for $L-n$ times of comparison. If you use the sliding window method, the basic idea is that every time you slide to the right, the new element that comes in on the right side of the window subtracts the old element that exits on the left, you get the sum of the previous sum and you get the new sum.

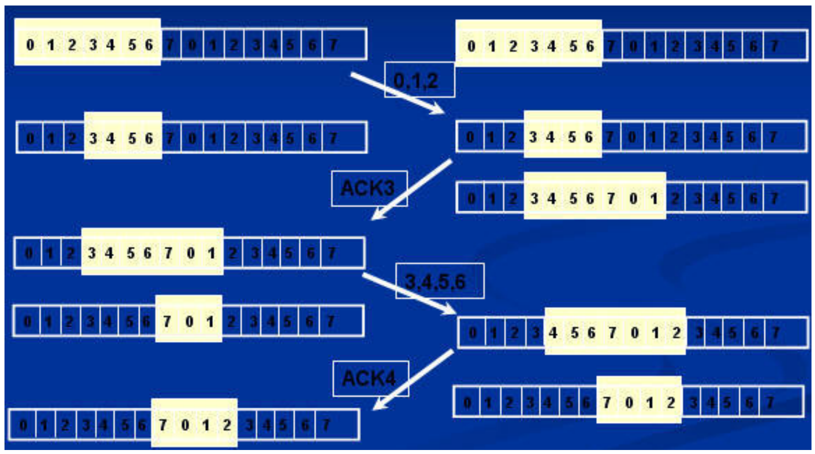

Figure 4.1: The schematic diagram of the Rolling model

It can be seen that the application scenario of sliding window has several characteristics: (a) the results that need to be output or compared are arranged continuously in the original data structure (the continuous string does not repeat the substring, and the continuous element in the array is the maximum sum); (b) every time the window slides, you only need to observe the changes of the elements at both ends of the window. No matter how long the window is, only two head and tail elements are operated each time. When the window used is relatively long, the number of operations can be significantly reduced; (c) the integrity of the elements in the window is relatively strong. Window sliding can be realized only by operating the change of the two positions of the head and tail, but all the elements in the window are often used to compare the results.

4.3.3 ARIMA model

The so-called ARIMA model refers to that the model is established by transforming the non-stationary time series into stationary time series, and then the dependent variable is returned to its lag value and the present value and lag value of the random error term. ARIMA model consists of moving average process (MA), autoregressive process (AR), autoregressive moving average process (ARMA) and ARIMA process according to whether the original sequence is stationary and the part contained in the regression.

The basic idea of the ARIMA model is to treat the data sequence and random sequence formed by the predicted objects changing with time and approximated by a certain mathematical model. Once identified, ARIMA model can work well to predict future values from past and present values of time series. To some extent, econometric models and modern statistical methods have been able to help enterprises predict the future.

The basic model of ARIMA prediction program. According to the graph of scatter, partial autocorrelation function and autocorrelation function graph of time series, the seasonal variation, trend and variance rule are tested by ADF unit root, and the stationarity of the sequence is identified. Generally speaking, the economic operation for time series of is not a stationary series. Carry out stabilization treatment for non-stationary sequences. If the data sequence is not of the stationary and has a certain trend of growth or decline, the data needs to be processed differentially; if the data has heteroscedasticity, and the data needs to be processed technically until the value of partial correlation function and autocorrelation function of the processed data are not significantly different from zero. According to the recognition rules in time series model, the corresponding model is built. If the partial correlation function of stationary sequence is truncated and the autocorrelation function is tailed, it can be concluded that the sequence is suitable for AR model. If the partial correlation function for a stationary sequence is tailed and the autocorrelation function is truncated, it can be concluded that the sequence is suitable for the model of MA. If the autocorrelation function and partial correlation function of stationary sequence are trailing, the sequence is well suitable for ARMA model [110].

In the end, parameter estimation will be carried out to test whether it has statistical significance, and the assumption test is carried out to diagnose if the residual sequence belongs to white noise; and the hypothesis test is carried out to diagnose if the residual sequence belongs to white noise.

The ARIMA model was implemented with the Python package 'statsmodels'. The best combination of (p, d, q) is determined by a grid search within the range of $[0, 10]$ for each parameter. The parameter combination with best Akaike information criterion (AIC) on the training set was chosen and the coefficients were fitted using OLS algorithm. The chosen parameters and fitted coefficients were used to forecast on the test set [111, 112]. For the ARIMA model, the best combination of (p, d, q) is $(2, 2, 4)$ through grid searching of AIC on the training set.

4.3.4 GM-LSTM model

The GM-LSTM model was implemented with the Python deep learning package 'PyTorch 1.0'. After the training of gray model, the residual fluctuation on the training set and sliced into fixed-length time series using the same rolling window of the gray model. This resulted in 22 samples for the model of LSTM training. The model of LSTM included one LSTM layer with 128 hidden neurons and one linear layer to learn the mapping between history time series [113, 114] and the residual fluctuation as described in Section 3.2.1.3. We use the Adam optimizer to update the weights for the LSTM network with a learning rate of 0.0005. In total, 200 epochs of training were performed as the training error converged.

As mentioned before, for the data of inbound tourist to Xi 'an from all countries along the OBOR , despite the large time span (29 years), but there are still insufficient amount of data, lack of categorical data, and high nonlinear characteristics. GM-LSTM model, integrates the advantage of GM that can predict the trend well and the characteristics of the LSTM model to calculate residual reasonable, so GM-LSTM model provides a good method to coupe with the problems of high-nonlinear degree and small amount of data in this study and gives a full analysis and explanation of the tourism prediction. In addition, the rolling window calculation method makes the prediction of the number of tourists to Xi 'an from all countries along the OBOR get a better result.

For convenient comparison in effectiveness of the two types of models in predicting tourists inbound to Xi 'an, we will use two types of data for analysis: first, we will compare the results of IEKF-RBF and IUKF-RBF, that is, the two results of the same model with different filtering methods in order to select the optimization model. Then it is compared with SDNM model and LSTM. Then, the traditional models are compared to choose a better model. Finally, the results of the most effective models of the above two types of models are selected and analyzed with the predicted results of GM-LSTM proposed in this thesis. The results and analysis will be discussed in next chapter.

Chapter 5

Results and discussion

In order to analyze the effectiveness of the GM-LSTM model proposed in this thesis, we select six models with good performance in tourism prediction and divide them into two groups: one group is the prediction model related to neural network, RBFNN with IEKF and IUKF, LSTM and SDNM; The other group is the classic traditional models: ARIMA, GM, Rolling GM. Generally speaking, the prediction results of the same problem with the same data set and the same simulation software have the same general trend, but the results may have certain differences. Therefore, for the convenience of research, the mean square error is used to evaluate the effectiveness of the models and screen the most suitable model. For the classical models ARIMA, GM and Rolling GM, we use the same idea to screen the most effective model. After the comparison model is selected, it is compared with the GM-LSTM model proposed in this thesis.

By analyzing the two sorts of models in the above chapters, this chapter will analyze and compare the calculated results. The calculation results of all models, including GM-LSTM model, are expressed as follows .

5.1 The results comparison of models related to neural network model

5.1.1 The results comparison of the RBFNN models with IEKF and IUKF

Firstly, we will choose the better one from the RBFNN models with IEKF and IUKF. According to the experimental results, we know that: (1) the suggested models can provide excellent forecasting accuracy with few and nonlinear tourist data; (2) although the both IEKF-based-RBF model and IUKF-based-RBF model have excellent real-time performance, the IEKF-based-RBF model has a better real-time performance than the IUKF-based-RBF model. And the main reasons of the obtained results is that: (a) the IUKF and the IEKF can effectively reduce the error of estimation for the state vector composed of the weight matrix $\boldsymbol{W}$ and the elements of the prototypes v_j by increasing the iterated number for the same value of the observation , so the accurate forecasting results can be obtained ; (b) the presented dynamic model is nonlinear according to the observation equation, but the equation of the state transition is linear. Therefore, the nonlinearity of the dynamic model may not be drastic enough and the IEKF-based-RBF model may obtain almost the same or even better accuracy prediction than the IUKF-based-RBF model while the iterative number reaches the given value.

Table 5.1: The forecasting results from 2014-2018 (Unit: ten thousand)

Year	Arrivals	IEKF-RBFNN	IUKF-RBFNN
2014	124.23	120.20	119.27
2015	110.72	109.10	107.78
2016	133.88	130.90	131.31
2017	175.13	176.19	175.80
2018	202.75	208.75	210.75

Table 5.2: Comparisons of AMSE for yearly tourist arrival forecasting

SMOOTHING	MAPE (%)
IEKF-RBF	24.82
IUKF-RBF	25.77

The data predicted by IEKF-RBF and IUKF-RBF models are shown in table 5.1. Table 5.2 shows the AMSE and running time. It can be seen that IEKF-RBF model is obviously superior to IUKF-RBF model in both AMSE and running time. Therefore, we compared the prediction results of IEKF-RBF model with those of other models discussed in the following.

5.1.2 The results comparison of models related to neural network: SDNM, LSTM and IEKF-RBFNN

For the well comparison of the validity of the forecast, we ignore the running time and only compare the error cases of the three models of SDNM, LSTM and IEKF-RBFNN, and the prediction results of the models related to the neural network are listed in table 5.3.

Table 5.3: Results of SDNM, LSTM and IEKF-RBFNN

model	SDNM	LSTM	IEKF-RBFNN
AMSE	27.09	22.33	24.82

WE can see from the data in table 5.3, although it is the same data set, the calculation results of different neural network models are quite different. The difference between LSTM and IEKF-RBFNN is small, while the difference between LSTM and IEKF-RBFNN is large, which also indicates that the different models have different results for the same data set. That is to say, the model is sensitive to data: for a given data set, there should be a model that fits it data set. So, for the problem we studied, we compare and study LSTM model with other traditional models and try to find out the best model to predict the Xi' an international tourist arrivals from the countries along OBOR. From the table 5.3, we can find that the LSTM model is more effective and accurate when three kinds of neural network-related models are applied to forecast the Xi' an international tourist arrivals from the all foreign countires. Therefore, We choose LSTM model to compare with traditional classical models.

5.2 The results comparison of traditional models

We used a simple comparison method to screen out the best model from the three models related to neural network, and then we will compare the three traditional models: ARIMA, GM and *Rolling* GM, and will screen the best effective model. For the well-comparison, we use three evaluation criteria: MSE, MAE and MAPE to compare the three traditional models, also known as classical models, to screen out the best model.

For comparison, we will firstly list the results of calculation of the classical models, and then analyze the performance of all models by graphs, and finally screen out the best model. The calculation results of the models of ARIMA, GM, Rolling GM are shown in table 5.4.

5.2.1 The results of ARIMA

Four limitations of the autoregressive model: the prediction of the model uses its own data, that is, the data used for modeling is the same set of data as the data used for prediction, the data used must be stationary, the data used must have autocorrelation, if the autocorrelation coefficient is less than 0.5, the autoregression model should not be used, and the autoregressive model is only suitable for predicting the phenomena related to the earlier period.

The results showed from Figure 5.1 to Figure 5.4 are the diagnostics plots of ARIMA$(2, 2, 4)$, and they are expressed as follows.

(a) The residual plot of ARIMA$(2, 2, 4)$ is showed in Fig 5.1., which the normalized error should fluctuate around zero. This figure shows that the fitting error of our adopted ARIMA$(2, 2, 4)$ is in a reasonable range. If the residuals are beyond a reasonable range, the reliability of the model is questionable. This graph proves the rationality of fitting from the aspect of fitting.

(b) The residual distribution plot of ARIMA$(2, 2, 4)$ is showed in Fig 5.2. The normalized error distribution should follow the normal distribution of N(0,1). The blue histogram and the orange core estimation line in the figure 5.2. have a high fitting degree with the theoretical normal distribution line in the green. It shows that the distribution of errors in the model ARIMA$(2, 2, 4)$ we used is approximate to the normal distribution, that the data meet the assumptions of the model we used, and that the model selection is reasonable.

(c) The residual quantile distribution plot of ARIMA$(2, 2, 4)$ is showed in Fig 5.3. This diagram is similar to the previous two plots: The abscissa is the theoretical quantile distribution of normal distribution, and the ordinate is the actual quantile distribution. The Fig 5.3. shows that they are uniformly distributed at both ends of the red line, indicating that the residuals satisfy the normal distribution with a mean of zero.

Table 5.4: The calculation results of the three models

Year	Arrivals	ARIMA	GM	Rolling GM
1980	4.00	0	4	*
1981	6.71	5.961690504	18.87833501	*
1982	9.09	8.869110917	19.98258485	*
1983	12.38	10.96578293	21.15142553	*
1984	15.13	14.74505756	22.38863518	*
1985	21.15	18.15956551	23.69821289	*
1986	25.78	25.40038446	25.08439169	*
1987	30.15	28.98072611	26.55165219	*
1988	36.58	32.97973801	28.10473712	*
1989	21.20	37.75353348	29.7486666	*
1990	25.88	30.76863914	31.48875438	*
1991	31.00	33.79535918	33.33062507	*
1992	40.16	40.06899	35.28023224	37.74875638
1993	43.50	50.95999669	37.34387772	41.69955665
1994	41.49	39.28969711	39.52823194	45.48264311
1995	41.35	40.42863314	41.84035553	46.5070237
1996	45.39	42.56674712	44.28772208	46.42471049
1997	48.53	51.30958098	46.87824237	48.35334168
1998	47.98	55.49383813	49.62028988	51.35593409
1999	55.41	52.86218865	52.52272788	53.55448406
2000	65.03	55.5004678	55.59493809	59.59715583
2001	67.20	63.51597563	58.846851	65.00381009
2002	74.13	66.65048624	62.28897794	69.50201209
2003	33.66	67.88308351	65.9324451	75.84953181
2004	65.03	51.21219448	69.78902947	62.54069271
2005	77.56	71.48821334	73.8711969	65.85812896
2006	86.73	83.9156242	78.19214242	72.82145819
2007	100.01	95.99605631	82.76583286	81.63230596
2008	63.20	66.75586895	87.60705203	94.79160375
2009	67.29	70.36137943	92.73144847	86.59586942
2010	84.18	77.57882938	98.15558607	81.03470657
2011	100.23	95.08703468	103.8969976	83.74470749
2012	115.35	117.9230296	109.9742413	93.93771394
2013	121.11	95.51135682	116.406961	109.3825155
		Test data		
2014	124.23	96.75275133	123.2159496	125.577586
2015	110.72	104.4490707	130.4232161	135.112301
2016	133.88	121.0634681	138.0520571	145.370958
2017	175.13	136.7592222	146.1271315	156.408523
2018	202.75	127.1896815	154.6745411	168.284136

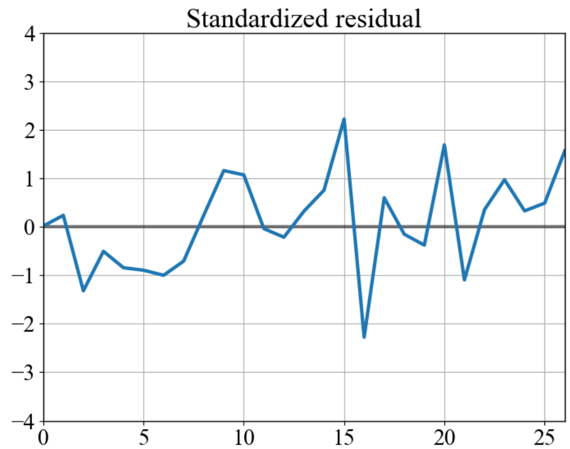

Figure 5.1: The standardized residual

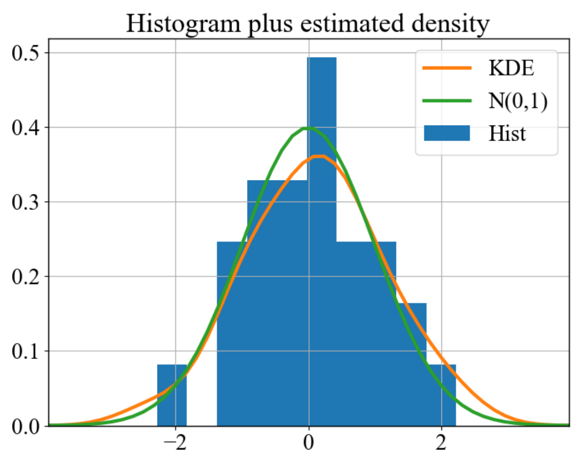

Figure 5.2: The histogram plus estimated density

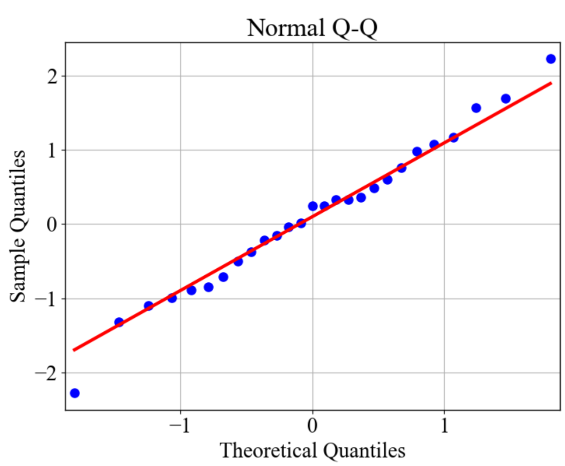

Figure 5.3: The plot of residual quantile distribution

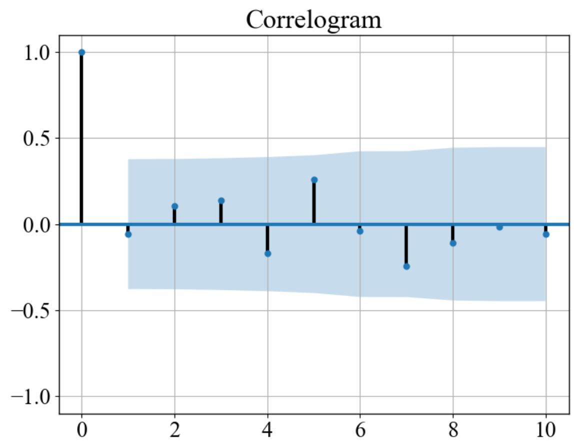

Figure 5.4: The plot of residual autocorrelation

(d) The plot of residual autocorrelation of ARIMA$(2, 2, 4)$ is showed in Fig 5.4. As can be seen from figure 5.4, all residual autocorrelation points fall within the blue confidence interval $(-0.5, 0.5)$, indicating that the parameter selection is reasonable.

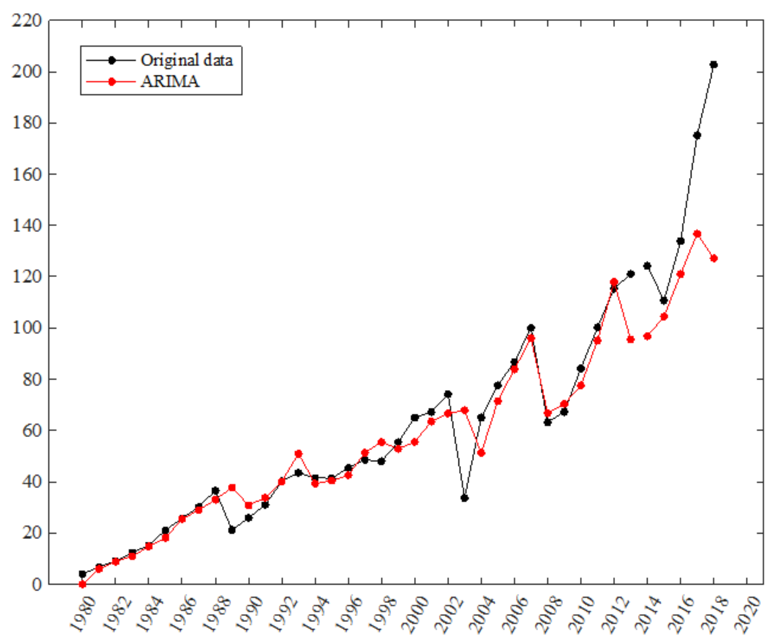

Figure 5.5: The fitting plot of ARIMA model

(e) The figure 5.6 indicates the data of ARIMA$(2, 2, 4)$ in the training set has a good fitting degree with the original data. It Illustrates the ARIMA$(2, 2, 4)$ model is effective. However, on the test set, the fitting degree is not very good, and even the trend of data has a big change, indicating that the model is not effective for the test set, so the accuracy of using ARIMA model to predict OBOR is not very good. It is very clear that the residuals tend become larger from figure 5.6 below. The results will be discussed later.

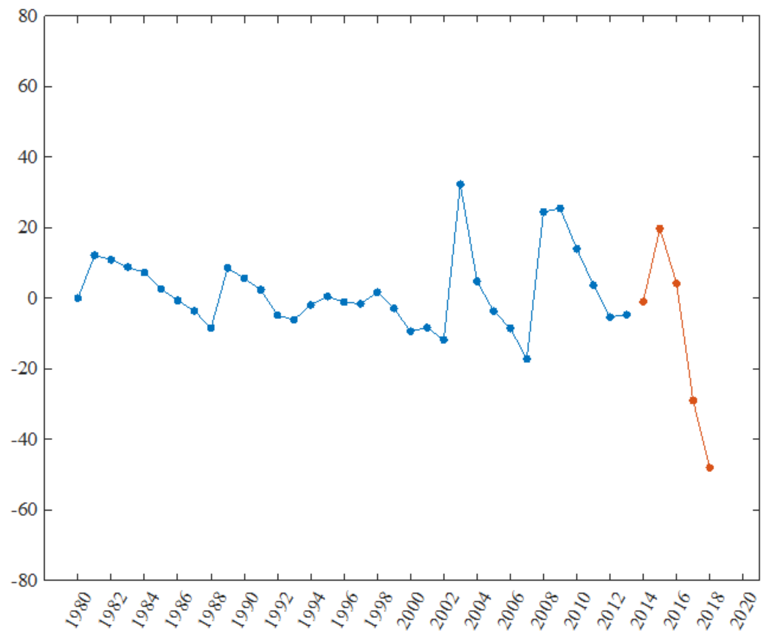

Figure 5.6: The residual error plot of ARIMA model

According to the above analysis, the model we selected is reasonable and the model parameters are appropriate for the data we studied. Therefore, we can believe that the calculated results of ARIMA$(2, 2, 4)$ are reliable. However, reasonable models and appropriate parameters do not always get the desired results, so the prediction of the tourists number of OBOR to Xi'an still needs to be analyzed and studied, and the ARIMA$(2, 2, 4)$ model, as a good performance classical model, can be used to compare with the GM-LSTM model proposed in this thesis.

5.2.2 The results of GM and Rolling GM

5.2.2.1 The results of GM

The model of GM$(1, 1)$ can fit on the whole training set and capture the overall trend well but can not include the fluctuation.

It can be seen from Fig.5.7 that the red line is the fitting result of the GM model. The fitting degree of both the test set and the training set is high. Therefore, compared with the model of ARIMA, the GM model is more effective in forecasting the development trend of tourist arrivals to Xi'an from the countries along the OBOR. However, similar to the model of ARIMA, the GM model test set has large residuals error, which is showed in Fig.5.8, (the red line is the residuals of the test set), despite the good fitting trend to the original data.

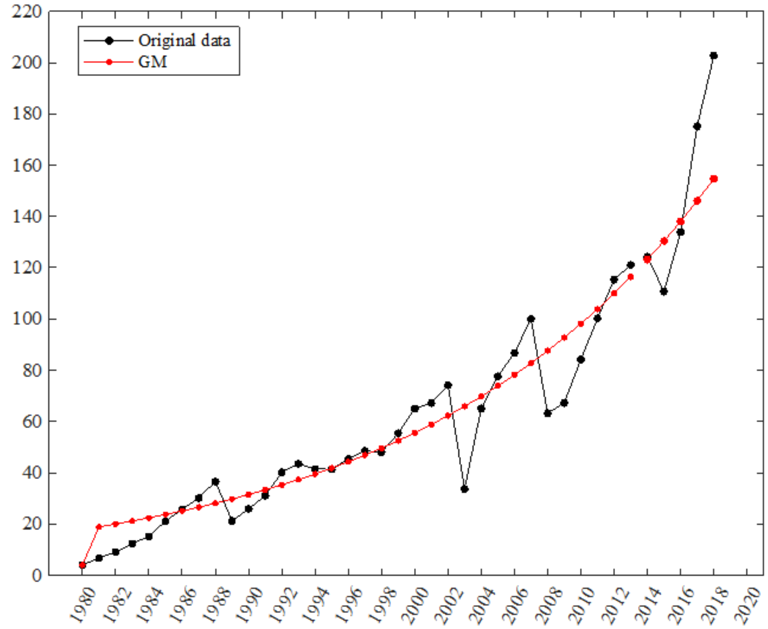

Figure 5.7: Comparison of the calculated results of the GM model with the original data

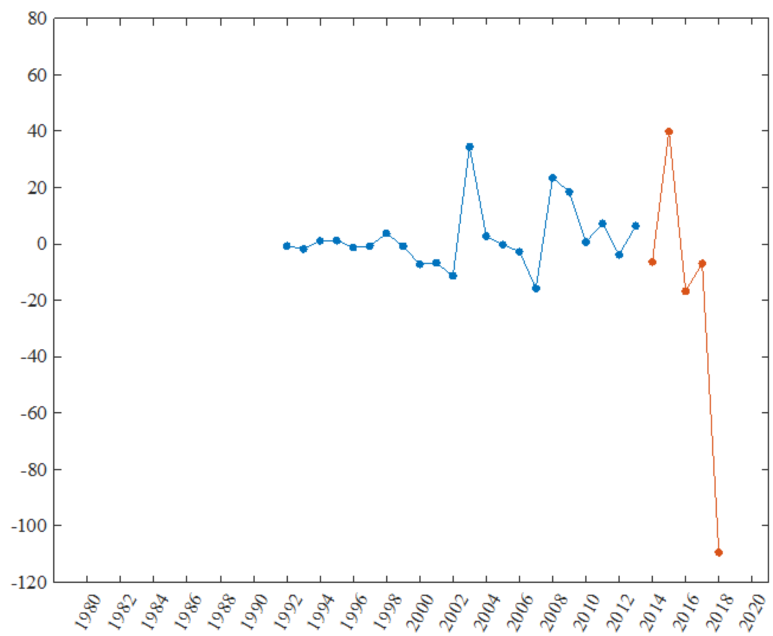

Figure 5.8: The residual error plot of GM model

5.2.2.2 The results of rolling GM

Determine the rolling window length. The length of the rolling window relies on the structure and characteristics of the data. In general, the difference is that the data set has different rolling window lengths. To determine the rolling window length of data set, we used MAPE performance as the criterion to determine the optimal rolling window length. The specific process is as follows:

For rolling $\mathrm{GM}(1,1)$ model, it is very necessary to determine the length of the window of the rolling GM model for comparison, therefore, we will compare the first-order MAPE performance of the training set with different rolling window lengths. The different rolling window lengths for $\mathrm{GM}(1,1)$ model is showed in Figure 5.9.

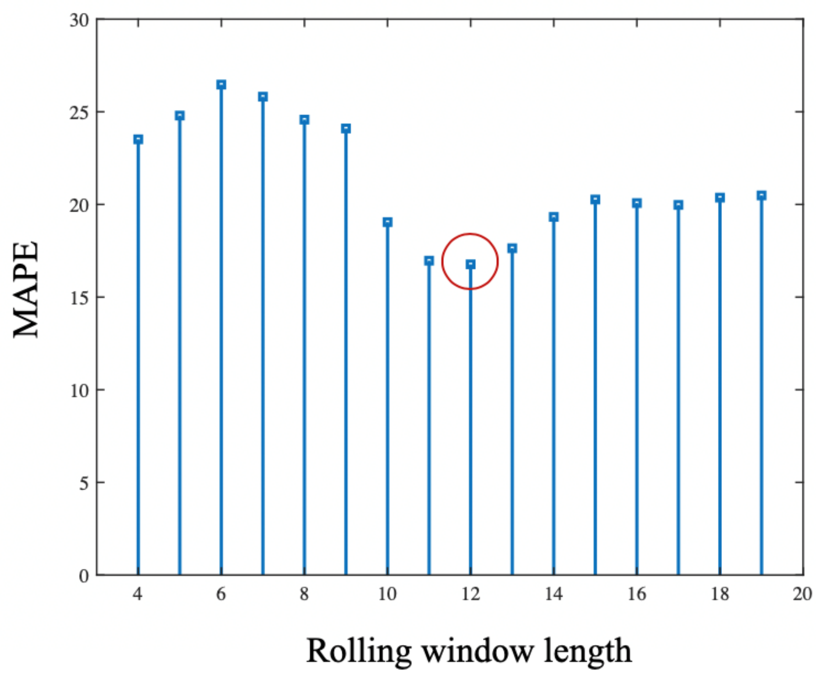

Figure 5.9: The length of rolling window and corresponding mean absolute percentage error (MAPE) of one-step-ahead prediction of rolling $\mathrm{GM}(1,1)$ on the training set

As shown from figure 5.9, there is a horizontal s-shaped distribution average absolute error to the length of the rolling window for one-step-ahead prediction. That is to say, the error of the early data is large, with the smallest in the middle, and the error in the back of the data is relatively stable. This error distribution indicates that the early data has a higher degree of nonlinearity and the later data has a lower degree of nonlinearity. Taking the minimum error as the criterion, the length of the window is set to 12, as shown in Fig. 5.9.

As can be seen from Fig. 5.10, since the length of the rolling window is 12, the fitting data is significantly reduced, but the fitting degree in the training set and test set is significantly higher than GM. Meanwhile, the variation trend of the rolling GM model is closer to the trend of the original data. Looking at figure 5.11, the residual error in the test set are significantly smaller than GM. It can be seen from the two figures that the rolling GM model is more effective than the GM model in studying the OBOR tourism problem. Specific quantitative analysis is discussed later.

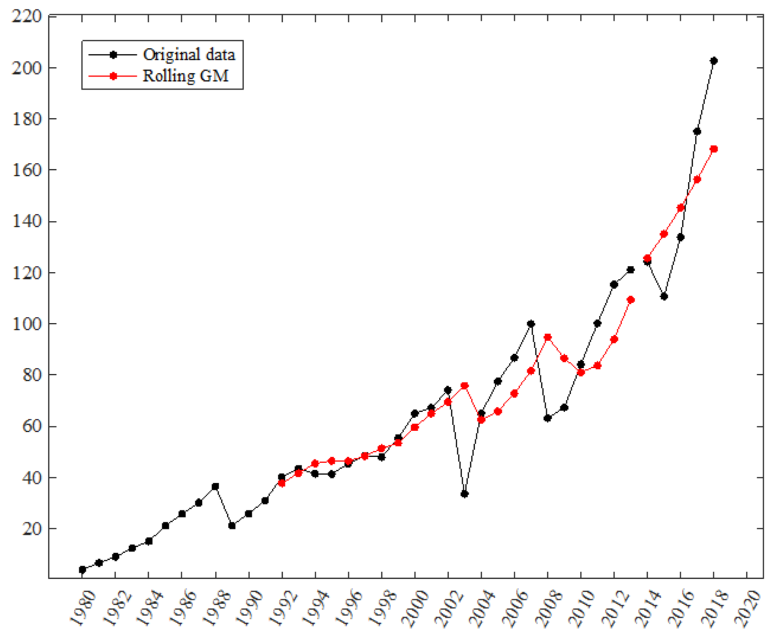

Figure 5.10: The comparison of the results: the rolling GM model and the original data

5.2.3 The results comparison of models: ARIMA, GM and the rolling GM

(a) Figure 5.12 is a line graph with the results of ARIMA$(2, 2, 4)$ and GM$(1, 1)$ and original data.

As can be seen from the figure 5.12, in the early data on the training set, the ARIMA$(2, 2, 4)$ model is obviously more effective than the GM$(1, 1)$ model, but with the increase of training data, the GM$(1, 1)$ model shows obvious stability. Looking at the test set, neither model performed very well, but the GM$(1, 1)$ model had better predictive performance than the ARIMA$(2, 2, 4)$ model. Since we focus on the final prediction result and the performance of the training set is not very important, we think that the prediction model of GM$(1, 1)$ for this data set performs better than that of ARIMA$(2, 2, 4)$, and we will compare the GM$(1, 1)$ model with the rolling window GM$(1, 1)$.

For the ARIMA model, the best combination of (p, d, q) is $(2, 2, 4)$ through grid searching of AIC on the training set. The diagnostics figures of ARIMA$(2, 2, 4)$ model are plotted in Figure 5.1, which shows a good fitting result. On the training set, the ARIMA$(2, 2, 4)$ model follows the trend and fluctuation well and obtained a small MAPE 15.73%. However, it fails to follow the original data on the test data and the forecasting error is large where the MAPE is 19.31%. This indicates that the ARIMA model could not generalize well with only small-sample observations. In comparison, the model fitted on the whole training set captures the overall trend well while not includes the fluctuation, resulting a higher MAPE of 25.57% on the training set. It is noted that the forecasting

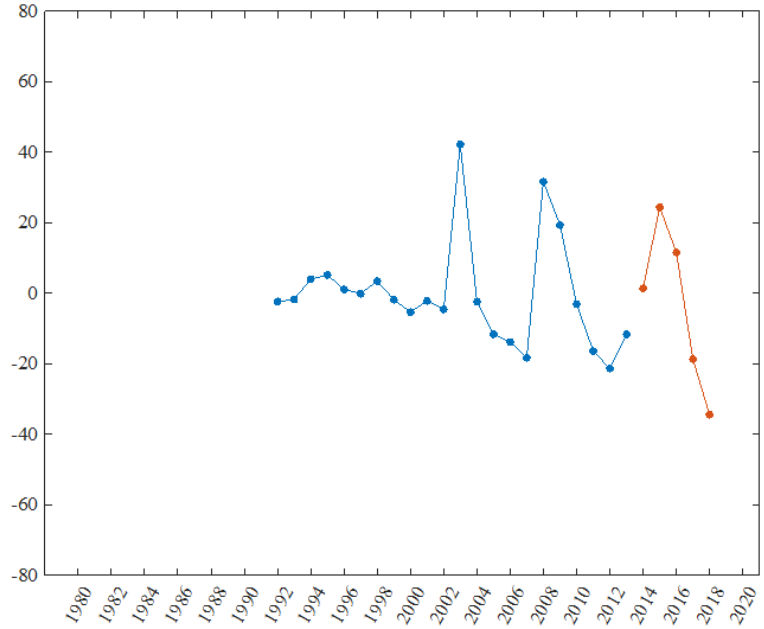

Figure 5.11: The residual error plot of rolling GM model

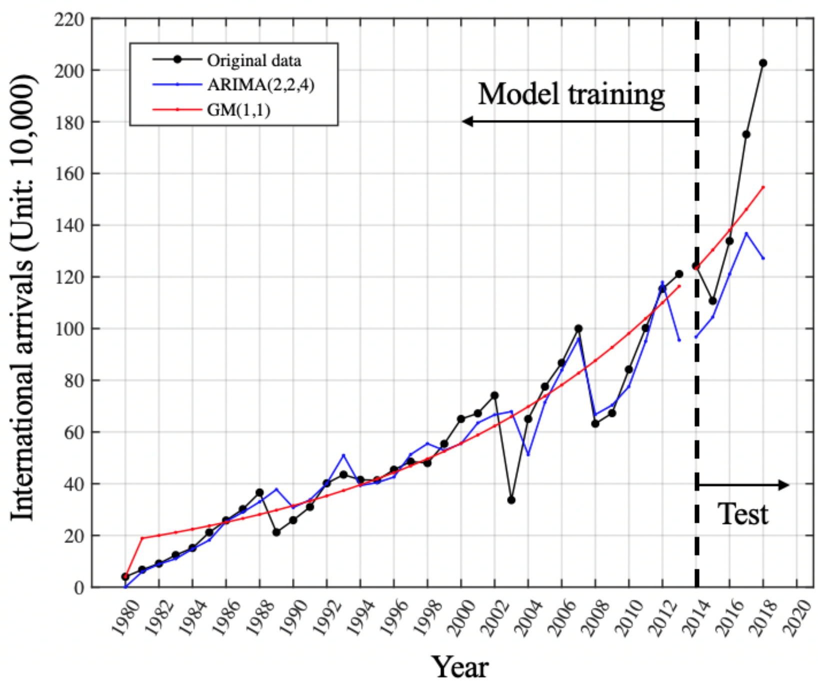

Figure 5.12: Comparison of ARIMA$(2, 2, 4)$ and GM$(1, 1)$ on the training and test set

error on the test set is 12.40% which suggests the good generalization ability of gray model based on limited data, as demonstrated in Figure 5.13.

(b) FIG.5.13 and FIG. 5.14 are the GM(1, 1) fitting diagrams with rolling GM(1, 1). Obviously, the fitting trend of rolling window is more close to the original data, the testing error of rolling window is smaller. So rolling window GM(1, 1) is better model for tourism prediction, and we will compare the rolling window GM(1, 1) model with the others.

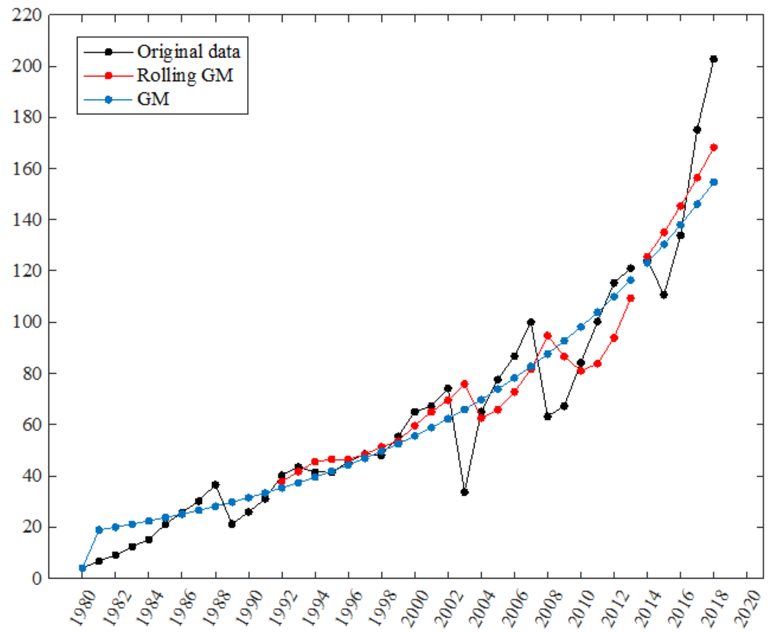

Figure 5.13: Comparison of rolling GM(1, 1) and GM(1, 1) on the training and test set.

For rolling GM(1, 1) model, we compared the MAPE performance of one-step-ahead prediction on the training set under different rolling window length. The results in Figure 5.9. suggests that the optimal window length for this time series is 12. Therefore, we trained a rolling GM(1, 1) with a 12-length window to capture the changing trend and forecast the latest five-year data on the test set (Figure 5.9). This adaptive model shows better performance than the classical GM(1, 1) model with the forecasting MAPE of 11.88% and captures a better increase tendency. However, from the one-step-ahead prediction results demonstrated in Figure 5.13, we observe that the rolling GM(1, 1) still cannot characterize the fluctuation well.

Now let's compare the validity of the roll window model. As can be seen from FIG. 5.7., their data in the training set are relatively close and the stability of the data is also good. However, in the test set data, the data of the rolling window model is more effective, closer to the real value, the prediction is more stable, and the effectiveness is higher. In the traditional model, we finally selected the rolling window model for further study.

From the above analysis, we can see that the rolling window-GM model is the best among the three classical models in prediction. Therefore, we screened out the rolling window-GM model from

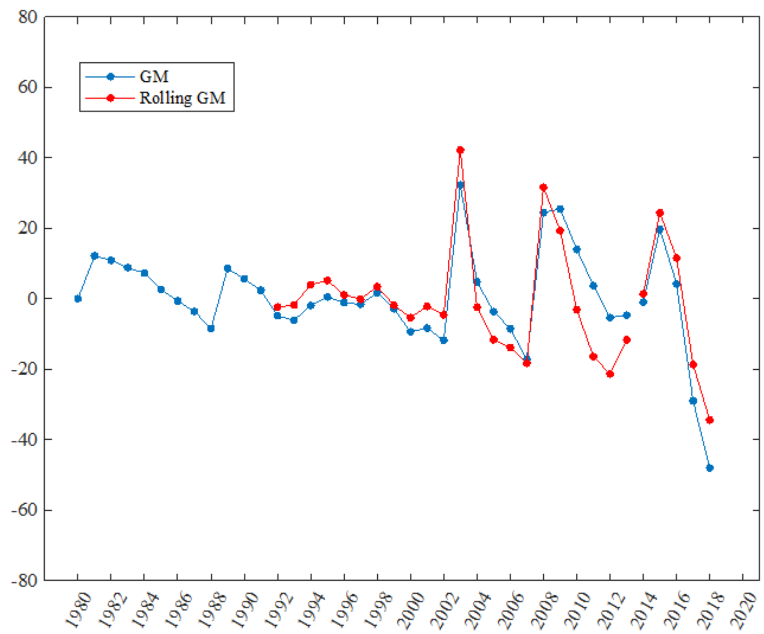

Figure 5.14: The residual error plot of rolling GM model and GM model

the three classical models to compare with the model proposed in this thesis.

5.3 GM-LSTM model

Through the above analysis and calculation, we selected the best two models: the rolling window-GM model and the LSTM model, which performed well in the tourism prediction from the six studied models. Now let's try to combine the advantages of these two models to generate the GM-LSTM model to predict the same problem with using the same data set and the same computing software.

5.3.1 The results of models

To analyze the effectiveness of the GM-LSTM model and facilitate the results in comparison, the results of the three type of model with the best performance: the rolling window-GM, LSTM and the GM-LSTM proposed are listed in table 5.5.

(a) The results of LSTM model The prediction of the number of tourists by LSTM model is not ideal from the data in the table 5.5. Similar to other models, the trend of the test set is quite inconsistent with the trend of the original data, the error is large, as shown in FIG.5.15 and FIG.5.16. So far, the two type of models studied in this paper have a common feature: the development trend of the test data is almost the same, but the accuracy of the prediction results are different. The quantitative analysis of errors is discussed later.

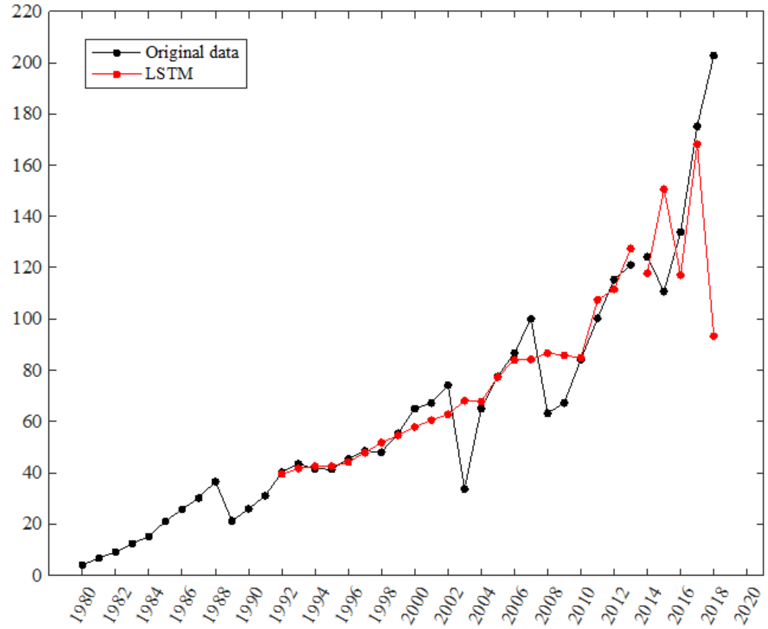

Figure 5.15: Comparison of the calculated results of the rolling LSTM model with the original data

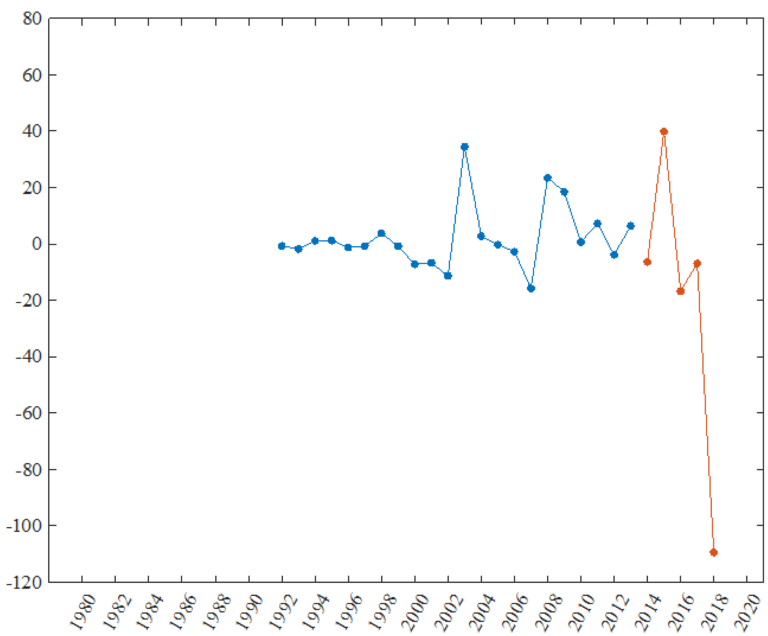

Figure 5.16: The residual error plot of LSTM model

(b) The results of GM-LSTM model As shown in table 5.3, among the three prediction models related to neural network selected in this paper, the model of LSTM has performed best among the neural network models in One Belt and One Road travel prediction. Therefore, we use this model to predict and modify the error between rolling-GM and the original data.

For the sake of clarity, the corresponding results are listed in table 5.5. Where, $E_{r\,\mathrm{GM}}$ is the error between the original data and the prediction results of rolling-GM model, while E_{LSTM} is the correction error calculated by LSTM model. The predicted result of GM-LSTM is that the result of rolling-GM model plus the E_{LSTM} .

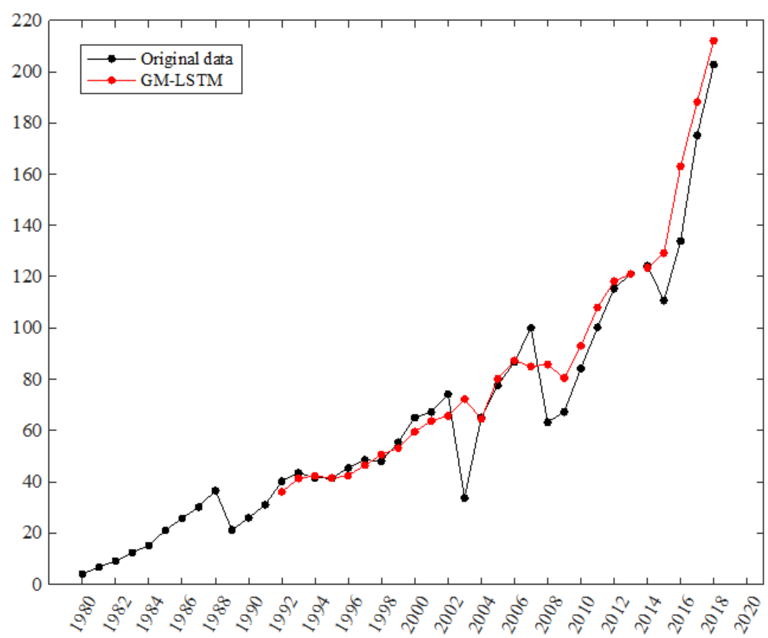

Figure 5.17: Comparison of the calculated results of the GM-LSTM model with the original data

It can be seen from figure 5.17 and figure 5.18 that the predicted results of the GM-LSTM model not only have the same development trend with the original data, but also have a very small error to the original data.

5.3.2 The result comparison of LSTM model, rolling GM model and GM-LSTM model

(a) The result comparison of rolling GM model and GM-LSTM model In FIG. 5.19, it is obvious that the GM-LSTM model is more effective than the rolling GM model in predicting the data in the test set, regardless of trend or error.

(b) The result comparison of three models Fig. 5.19 shows the original data (black), the rolling GM model (blue line), the LSTM model (green line), and the fitting curve of the GM-LSTM model (red line) proposed in this paper. Since the length of the window is 12, the data fitting starts

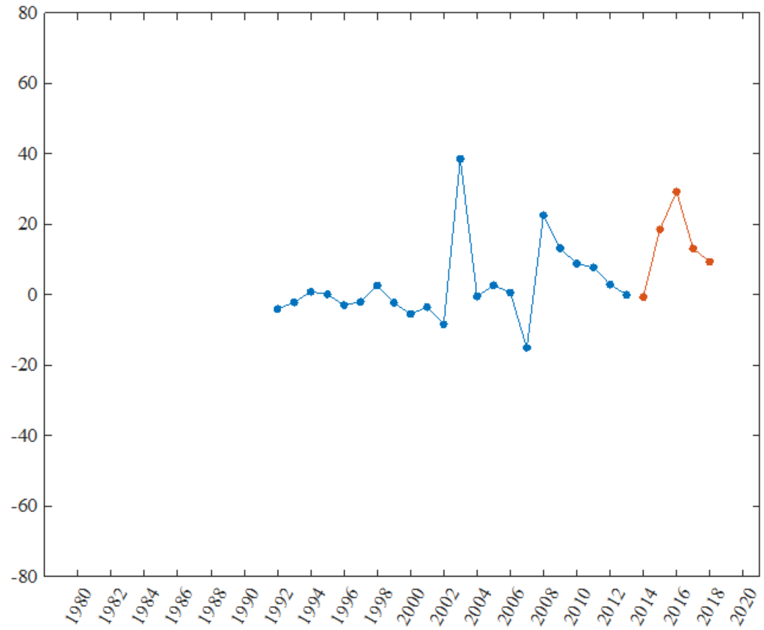

Figure 5.18: The residual error plot of GM-LSTM model

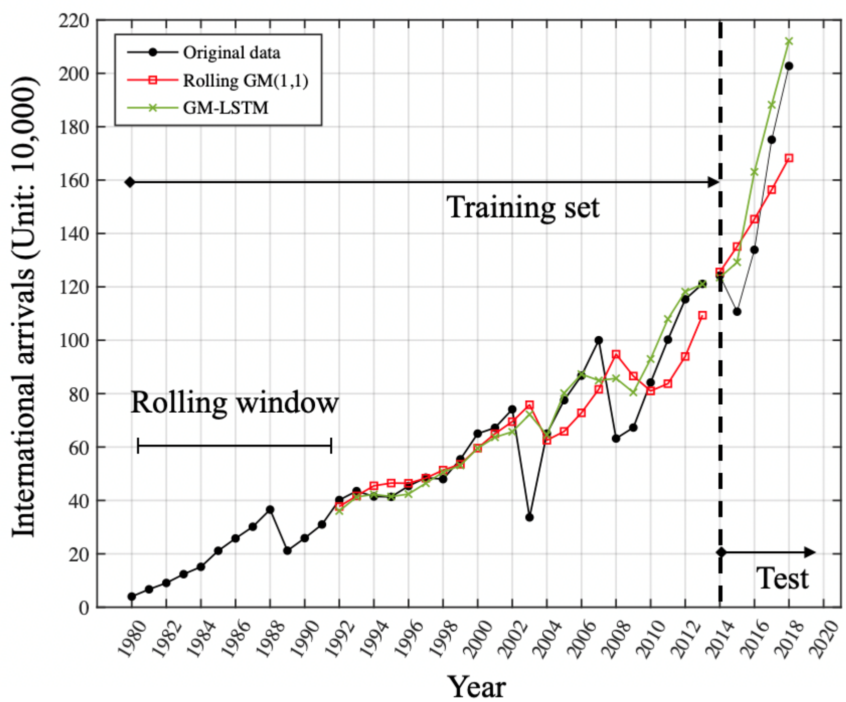

Figure 5.19: Comparison of the calculated results of the GM-LSTM model with the original data

from the data in 1992.

Now, we focus only on the predicted results of the test set. Obviously, from the development trend of the three model's test sets, the GM-LSTM model has the best performance in prediction effect, the rolling GM model is the medium, and the LSTM model has a poor prediction effect. As can be seen from Fig. 5.21, the error values of the three model test sets are as follows: the prediction result error of the GM-LSTM model is the smallest, the error of the rolling GM model is medium, and the result error of the LSTM model in prediction is the largest among three models.

The data showed in the table 5.6. are not all result data because the rolling window length is 12. For rolling GM(1, 1) model and GM(1, 1)-LSTM model, there are only 12 data in training set from 1992 to 2013, for the comparison, the corresponding 12 data in the LSTM model are taken, but all of the test set are the same. Figure 5.9 is the line graph of the above three model training sets and test sets.

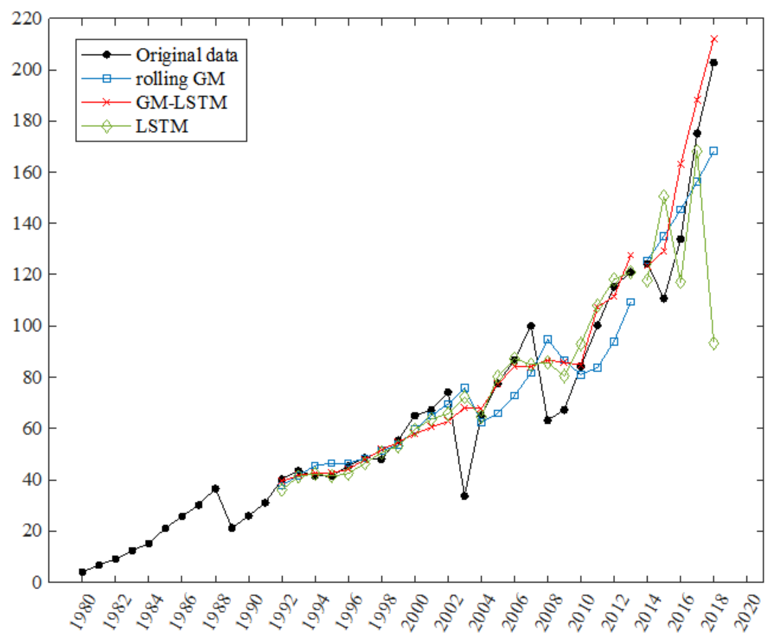

Figure 5.20: Comparison of the GM-LSTM, rolling GM, and LSTM model with the original data

FIG. 5.20 and FIG. 5.21 fully illustrate the effectiveness of the GM-LSTM model in predicting the number of tourists arrivals in Xi'an from the countries along the route of One Belt and One Road.

The curves of the three models can be shown from FIG. 5.19. Obviously, the GM-LSTM model has the best effectiveness, followed by rolling GM, and then the LSTM model. Although the validity of the three models can be seen from the figure, in order to further illustrate the validity of the three models, we still use the three test error standards mentioned in the previous chapter to make a comparison, and the result is listed in table 5.6.

FIG. 5.20 is the error in the test set of three models. And the error of the GM-LSTM model is the smallest, the error of the rolling-GM is medium, and the error of the LSTM model is the

Table 5.5: The calculating results of the models of rolling window-GM, ELSTM and the GM-LSTM

Year	ErGM	ELSTM	Rolling GM	GM-LSTM
1980	*	*	*	*
1981	*	*	*	*
1982	*	*	*	*
1983	*	*	*	*
1984	*	*	*	*
1985	*	*	*	*
1986	*	*	*	*
1987	*	*	*	*
1988	*	*	*	*
1989	*	*	*	*
1990	*	*	*	*
1991	*	*	*	*
1992	2.41124362	-1.685789097	37.74875638	36.06296728
1993	1.80044335	-0.400096533	41.69955665	41.29946012
1994	-3.99264311	-3.191798244	45.48264311	42.29084487
1995	-5.1570237	-5.053766937	46.5070237	41.45325676
1996	-1.03471049	-4.039922627	46.42471049	42.38478786
1997	0.17665832	-1.895284245	48.35334168	46.45805744
1998	-3.37593409	-0.779436076	51.35593409	50.57649802
1999	1.85551594	-0.434655896	53.55448406	53.11982816
2000	*	-0.061025605	59.59715583	59.53613023
2001	*	-1.343149549	65.00381009	63.66066054
2002	*	-3.798814519	69.50201209	65.70319757
2003	*	-3.617063387	75.84953181	72.23246842
2004	*	2.004575763	62.54069271	64.54526847
2005	*	14.33135594	65.85812896	80.1894849
2006	*	14.48988678	72.82145819	87.31134497
2007	*	3.302769332	81.63230596	84.93507529
2008	*	-9.054243934	94.79160375	85.73735981
2009	*	-6.121819589	86.59586942	80.47404983
2010	*	11.94449734	81.03470657	92.97920391
2011	*	24.22309348	83.74470749	107.967801
2012	*	24.26197953	93.93771394	118.1996935
2013	*	11.64013386	109.3825155	121.0226493
		Test data		
2014	*	-2.078577851	125.577586	123.4990081
2015	*	-5.881226262	135.112301	129.2310747
2016	*	17.69824469	145.370958	163.0692027
2017	*	31.79807964	156.408523	188.2066026
2018	*	43.79459512	168.284136	212.0787311

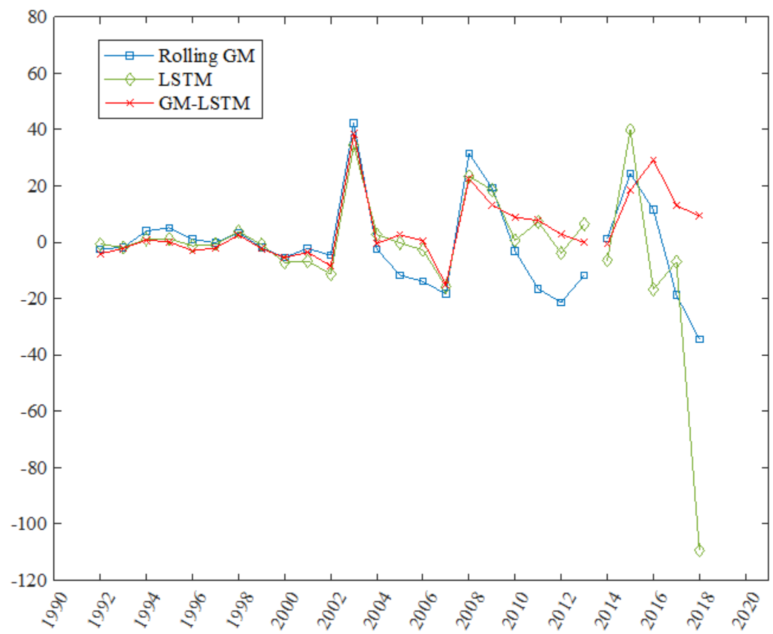

Figure 5.21: Comparison of the calculated results of the errors

largest. On the one hand, it shows the effectiveness of the GM-LSTM model, on the other hand, it shows that the effect of a single model to predict the travel problem with few data and high degree of nonlinearity is poor.

5.4 Quantitative analysis of three models

In order to fully illustrate the effectiveness and the performance of each model, we illustrate the validity of the three traditional models using the three criteria proposed in the previous chapter to verify their validity. The results are listed in table 5.6.

we will make a quantitative analysis of the characteristics and effectiveness of each model as follows:

As the previous research ideas, we do not pay attention to the performance of the model training set, and only analyze the data of the test set. The discriminant criteria were MSE, MAE and MAPE. The error values of three criteria for the two types of models are listed in table 5.6. According to the previous research, the classical models control the variation trend of the test set. It can be seen from table 5.6 that the LSTM model has the largest error, and the trend prediction of the change of the test set data is unreasonable for the LSTM model, which can only be used for residual prediction. However, the GM$(1,1)$ model and Rolling GM model in the classical model have a small prediction error, which is suitable for the trend prediction of test set data. Taking MAPE standard as an example, the MAPE values of the GM$(1,1)$ model and the rolling GM model are 12.40 and 10.88, respectively and they all fit the trend prediction of the test set data. In spite of the difference between the values of 12.40 and 10.88 is not big, but it is very meaningful to predict the tourism number of One Belt and One Road, because the predicted data unit in the table 5.6 is $10,000$. Therefore, any improvement and effectiveness of the prediction model is of great significance, because tourism is a comprehensive industry, which is closely related to the tourism layout and management of transportation, hotels and attractions. Since the accuracy of the prediction plays a crucial role in the city's tourism decision-making, the validity of the Rolling GM model in predicting the trend is more significant.

Similarly, the MAPE value of the GM-LSTM model is even smaller than that of the Rolling GM model, which is 9.24, indicating the prediction efficiency of the GM-LSTM model is the best.

Now, let's analyze the accuracy of the prediction, that is, the effectiveness of the model. It is very important for time series to predict the validity of tourism model. We take the number of tourists in the test set in 2018 as an example to analyze the prediction accuracy. As indicated in table 5.5, the error between the predicted value of GM-LSTM and the real value is only 93,328 thousand, while the error of the predicted value of Rolling GM model is 344,617 thousand. However, the number of tourists arriving in xi'an is concentrated in the summer from mid-Jjune to mid-September. If the error of the number of arrivals is too large, the significance of the prediction will be lost. As can be seen from the data in table 5.6, even if the difference of MAPE between the Rolling GM model and the GM-LSTM model is only 1.64%, the error of the corresponding predicted number of tourists is 33,25 thousand and for Xi'an, a tourist city, the error of arrivals within three months is relatively large. Therefore, the accuracy of prediction has practical significance for infrastructure construction, scenic spot management, tourist management, transportation, hotel and dormitory administrators.

Table 5.6: The calculating results of the models of MSE, MAE and MAPE

Models	Training set (1980-2013)			Test set (2014-2018)		
	MSE	MAE	MAPE (%)	MSE	MAE	MAPE (%)
ARIMA(2,2,4)	83.10	5.70	15.73	1267.95	32.10	19.31
GM(1,1)	115.55	7.80	25.57	711.69	20.39	12.40
Rolling GM	219.36	10.20	16.77	443.34	16.08	10.88
LSTM	120.99	8.11	12.99	2786.00	35.88	22.33
GM-LSTM	123.23	6.69	12.43	280.65	13.17	9.24

For the GM-LSTM hybrid model, we observe that the LSTM can help to characterize the fluctuation well on the training set. The GM-LSTM provides the best performance on both one-step-ahead prediction on the training set as well as the five-year prediction on the test set. This suggests the accuracy and efficiency of our proposed hybrid AI model which combines the strength of gray model and neural networks.

The detailed performance comparison of different models is shown Table 5.6. It is observed the proposed GM-LSTM hybrid model outperformed all other models in terms of both MSE, MAE and MAPE, in the forecasting task of Xi'an annual international tourist arrivals.

From the above research, in the prediction of tourism, the traditional models and the models related to neural network have their own advantages and characteristics, combine their advantages, in the study of tourism forecast, is a very good method in some degree, and it can even worth studying the model and algorithm.

In order to further study the validity of the related combination model, we also analyze several other combinations. The results show that among several combinations of classical models and neural network models, the GM-LSTM hybrid model is the most effective in predicting the number of tourists arriving in Xi 'an.

Chapter 6

Conclusion

Based on full study on the tourism models, there are two sorts of the artificial intelligence models in tourism prediction: classical prediction models and neural network models, and we screened out two good performance, high efficiency models: one is the classical forecasting model which is the rolling GM model and another one is LSTM model which is neural network forecasting model from the two types of models using the same calculation software, the same training sets and test sets, as well as the same evaluation standard, and then put the two perfect models together to build a new model. In the process of the research, the basic trend of the prediction of the tourism prediction problem, whether they are the classical models or the neural network models, is consistent with that of the same kind of problem, but the accuracy of the prediction varies greatly. So it is very importance and necessity to find the model which can accurately predict tourism problems.

In this thesis, aiming at the problem of less observations and nonlinearity in tourism demand prediction, we selected some prediction models with better performance in the prediction. On the basis of full comparison and analysis, we proposed GM-LSTM hybrid model which has better prediction accuracy of tourism demand than that of the selected model under the condition of less and nonlinear tourism data.

The contributions of this thesis are as follows: (1) Compared with other models, the GM-LSTM hybrid model we proposed is more effective. (2) The GM-LSTM model integrates the grey model (GM)'s advantages and the neural network (NN)'s advantages to make the observation based on small samples. (3) The GM-LSTM hybrid model is a good solution to the less of the problem in observations and nonlinearity in tourism demand prediction. (4) The forecasting accuracy of the GM-LSTM hybrid model is high, and it has great application potential in solving similar problems of time series prediction. (5) The GM-LSTM hybrid model also adapts to the prediction problem of large observation, the GM-LSTM hybrid mode therefore,can be used to increase the data set and complexity.

The model has the following limitations: (1) For the prediction methods, the limitation of current research is to establish a univariate time series prediction model, which has a weak explanatory power for tourists' behavior; (2) The adaptability of the model is not clear. In the future research, it is very necessary to verify the effectiveness of the model for multivariate prediction, and then carry out the multivariate analysis to have a deeper understanding of the prediction of tourists' behavior.

Prediction research is basically the forecasting of the future trend of the event and the quantitative prediction of the event based on the historical event's data . At present, the prediction model is successful in predicting the trend of events, but the accuracy of the quantitative prediction of events is far from ideal. Therefore, the forecasting results for time series are all studies on the regression value of events, trying to make the regression value close to the real value, but it cannot be accurate, let alone precise. The challenge for prediction model research is: firstly, how to make the prediction accurate, and secondly, how to describe the events with small probability that have a big influence on the prediction result in the model , such as epidemic disease, natural disaster (wind disaster, earthquake, extreme climate, etc.), and political factors. The research of this thesis mainly tries to solve the first challenge: to make accurate prediction.

According to the characteristics of each tourism data set, there should be at least one appropriate model for the prediction of tourism problems, so in the future research, finding the appropriate model is the direction we have been working on, so is true for other forecasting problems.

Chapter 7

Outlook and plan

Under the careful guidance of my doctoral supervisors, the main research work of this paper is as follows:

The number of inbound tourists to Xi 'an was studied and predicted macroscopically. At the same time, taking a scenic spot in Xi 'an as an example, the forecast of short-term tourist flow of scenic spot is analyzed and predicted systematically. Although more detailed and specific studies have been made on these two issues, satisfactory results have been obtained, there are still many things to analyze on related issues, the research methods are still worth exploring, and the research content is still worth expanding. The specific limitations of this research are as follows:

The acquisition of data resources in this research is limited and insufficient, with dimensions to be improved and data volume to be expanded. In particular, there is a lack of scientific analysis on the classification of national per capita income level, classification of cultural customs, and summary of travel characteristics in the source countries, as well as a lack of research on the acquisition methods and approaches of the above multidimensional data, as well as insufficient analysis and summary of the characteristics of tourism demand.

At the same time, there is no data analysis on tourism destinations, ignoring the relationship between Xi'an's tourism industry and various industries under the background of pan-tourism format, and lacking of data collection and analysis on the attraction of the formed industrial clusters to tourists. This study neither makes in-depth population prediction based on the data of the six elements of tourism, nor makes specific research on the cultural customs, travel characteristics and other important elements of countries along the route of OBOR .

The resource view of pan-tourism era is a dynamic industrial resource view. However, the selection of research objects in this study is not representative enough. As a representative of Chinese ancient civilization and culture, the outstanding tourism attraction elements of Xi'an are humanistic feelings and cultural heritage, which are relatively single and special. With the complexity and diversification of market in tourism demand and the blurring of tourism boundary, the research directions need to be expanded.

In this study, only one scenic spot is selected as the object of study, without in-depth research and exploration on the nature, cultural background and characteristics of urban scenic spots. We should pay more attention to the internal connection between different scenic spots, study more appropriate algorithms, plan more appropriate paths, and develop more reasonable tourism products.

Combined with city tourist arrivals forecast and specific spots of passenger volume forecast, the more appropriate calculation method and the research methods should be explored to dig out more valuable information, in order to making and providing theoretical basis for tourism infrastructure construction and government decision and providing more scientific and accurate decision support for the improvement of OBOR tourism service quality. For example, we should provide differentiated tourism products for different tourists, allocate tourism resources reasonably, and guide tourism behaviors by the country characteristics and regional differences of inbound tourists as well as the preference data of specific tourist attractions.

With the development in tourism industry and the accumulation of tourism data, the future research contents and plans are as follows:

According to different national characteristics and cultural backgrounds, to allocate tourism resources and formulate the appropriate tourism products by predicting the number of the top 10 countries with tourist arrivals.

To predict the total number of visitors from each country, then predict the development trend of tourist arrivals more accurately.

A reasonable and scientific model should be established to provide theoretical support and scientific decision-making for Chinese citizens to choose travel products.

Different analysis methods should be studied according to different characteristics and properties of scenic spots, and management strategies should be put forward to provide the theoretical basis for the government to make correct decisions.

Analyze and mine deeply on the existing data of various scenic spots, and study more reasonable algorithms to provide reasonable high-grade tourism products for international tourists.

Consider comprehensively the total arrivals of tourists in Xi'an and the number of scenic spots, study the corresponding algorithm, and put forward valuable Suggestions to fully develop the tourism resources of Xi'an.

The number of tourists from key countries should be predicted month by month, so as to grasp the change law of tourists and the total number of tourists more accurately.

Study China's preference for tourism products of countries along the OBOR as a tourism source country, and the impact of Chinese tourists' travel on China's cultural output.

Make use of two-dimensional data or multidimensional data to predict and study the same problem, compare the their results and draw meaningful conclusions.

In a word, it is necessary to research deeply on artificial intelligence methods, fully explore the connotation and cultural elements of tourist attractions, improve the grade and level of tourism products, so that more international tourists can understand the essence of ancient Chinese culture, and spread Chinese culture to the world, so as to enhance China's international influence, make our ancient civilization influence the whole and promote the development of world civilization and the construction of world culture.

Bibliography

[1] The New Trend, New Exploration and New Formats of Cultural and Tourism Integration April 22, 2019 15:52 Source: Economic Daily - China Economic Net

[2] Zhang qian: Pioneer of the Silk Road. CCTV. 2017-05-23.

[3] Bai changhong. Industry Talent Training under the Background of Cultural and Tourism Integration–Practical Demand and Theoretical Topic. People's Forum · Academic Frontier, June, 2019

[4] Yang jinsong, "One Belt And One Road" Cultural and Tourism Integration Has Broad Prospects, China Tourism News, 2018-4-13

[5] Zhang qian, Pioneer of the Silk Road. CCTV. 2017-05-23

[6] The Concept of China-central Asia New Silk Road Economic Development Belt. Central Asia Research Network. 2013-09-09.

[7] Li qiyuan, Empirical Research on the Impact of Foreign Exchange Income from Tourism on Economic Growth, Research on Financial and Economic Issues, No. 09, 2014.

[8] Zhou caifeng and Ren wenbo, Probe into the Hidden Worries Behind the Rise of Tourism in Countries along the "One Belt And One Road", [J]. Modern marketing, 2019(6): 8.

[9] Shaanxi tourism blue book, 2014, shaanxi tourism press: 22.

[10] Internal Information: 42.

[11] Victor Maier Schoenberg. Era of Big Data [M]. Hangzhou: Zhejiang People's Publishing house,2013: 67.

[12] Application Channels. The large data of shallow data structure and management [EB/OL]. [2012-09-10]. http://www.50cnnet.com/html/2012/dashuju_0910/33580.html

[13] Guo xin. Tourism big data and mining analysis research [J]. Computer knowledge and technology,2013,9(14):3215-3216.

[14] Lu yuan. Analysis of intelligent tourism mode from the perspective of big data [J]. Holiday tourism,2018(12):98-99.

[15] Zhang xiaohua, Guo xuan, Li Juan, Ma hao. Intelligent tourism information analysis system based on big data platform [J]. Fujian computer,2015,31(08):93-94.

[16] Tian qiuyang. Application analysis of big data in tourism management [J]. Holiday tourism,2018(12):100-105.

[17] Xiao jie. Application analysis of big data in tourism management [J]. Holiday tourism,2018(12):104-105.

[18] Zhang lijun, Zhao xia. Tourism management service system based on big data analysis [J]. Information communication,2014(11):232-233.

[19] Liu xiaoyan, Zhang min. Tourism management information system based on artificial intelligence [J]. Automation and instrumentation,2016(08):147-148.

[20] Meng zhihui, Liu guanglu, Hao chengyu. Analysis on the development path of tourist attractions based on the analysis of visitor flow data [J]. World of labor security,2017(36):63.

[21] Sun yan-ping, Zhang lin, Lu ren-yi. Neural network method for tourism source prediction [J]. Human geography,2002(06):50-52.

[22] Yu mingtao, Ye xiaotong. Improved BP neural network based on particle swarm optimization [J]. Microcomputers and applications,2015,34(21):51-54.

[23] Jing le, Development of China's tourism service trade under the background of "One Belt And One Road" [J]. Reform and strategy,2017,33(07):172-172

[24] Cheng qian, Development of tourism service trade in China under the background of "One Belt And One Road" [J]. Modern economic information,2017(10):140.

[25] Wang zhan-long. Influence of "One Belt And One Road" on tourism [J]. Tourism overview (second half),2018(12):46.

[26] Han zhiyong. The impact of "One Belt And One Road" strategy on asean tourism [J]. Rural economy and science and technology,2016,27(06):73

[27] Song hongjuan, jiang jade-shi. Value judgment of "One Belt And One Road" tourism market [J]. Development research,2017(04):149-155.

[28] Aslihan Dursun A M C. Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis [J]. Tourism Management Perspectives, 2016 (18) : 153 - 160.

[29] Jehn - Yih Wong H C P C. Identifying Valuable Travelers and Their Next Foreign Destination by the Application of Data Mining Techniques [J]. Asia Pacific Journal of Tourism Research, 2006, 11 (4) : 355 - 373.

[30] Yang ting. Development path of xi'an cultural tourism industry under the background of "One Belt And One Road" [N]. Xi'an daily,2019-06-25(007).

[31] Su hongxia, Zhang jie. Dynamic evolution characteristics of shaanxi international tourist market under the background of One Belt And One Road – based on statistical data from 2007 to 2016 [J]. Henan science,2019,37(04):684-688.

[32] Commentator of this newspaper. Build "One Belt And One Road" into the road of civilization [N]. China culture daily,2019-04-26(001).

[33] Wen ke, nearly 150 million outbound tourists visited China in 2018,China consumer news - China consumer network 2019-02-18.

[34] María Henar, Salas-Olmedo, Borja, Moya-Gómez, Juan Carlos,García-Palomares, Javier Gutiérrez. Tourists' digital footprint in cities: Comparing Big Data sources[J]. Tourism Management, 2018,66.

[35] Sheng-Hshiung Tsaur, Yi-Chang Chiu, Chung-Huei Huang. Determinants of guest loyalty to international tourist hotels—a neural network approach[J]. Tourism Management, 2002, 23(4).

[36] Stephen F. Witt, Lindsay W. Turner. Trends and Forecasts for Inbound Tourism to China[J]. Journal of Travel & Tourism Marketing, 2003, 13(1-2).

[37] Alfonso Palmer,Juan José Montaño,Albert Sesé. Designing an artificial neural network for forecasting tourism time series[J]. Tourism Management, 2005, 27(5)

[38] Chi Kin Chan, Stephen F. Witt,Y.C.E. Lee,H. Song. Tourism forecast combination using the CUSUM technique[J]. Tourism Management, 2009, 31(6).

[39] Jamal Shahrabi,Esmaeil Hadavandi,Shahrokh Asadi. Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series[J]. Knowledge-Based Systems, 2013, 43.

[40] Oscar Claveria, Salvador Torra. Forecasting tourism demand to Catalonia: Neural networks vs. time series models[J]. Economic Modelling, 2014, 36.

[41] Yen-Hsien Lee, Ya-Ling Huang. Accurately Forecasting Model for the Stochastic Volatility Data in Tourism Demand[J]. Modern Economy, 2011, 02(05).

[42] Haiyan Song,Stephen F. Witt. Forecasting international tourist flows to Macau[J]. Tourism Management, 2006, 27(2).

[43] Wei CHEN, Jian SUN, Nonmembers, Shangce GAO, Member, Jiu-Jun CHENG, Jiahai WANG and Yuki TODO'Using a Single Dendritic Neuron to Forecast Tourist Arrivals to Japan', IEICE TRANS. INF. & SYST., VOL.E100–D, NO.1 JANUARY, pp.190–202, 2017.

[44] Athanasopoulos, G., Hyndman, R. J., Song, H. Y., & Wu, D. C. (2011). The tourism forecasting competition. International Journal of Forecasting, 27(3), 822–844.

[45] Jungmittag, A. (2016). Combination of forecasts across estimation windows: An application to air travel demand. Journal of Forecasting, 35(4), 373–380.

[46] Liang, Y. H. (2014). Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan. Computers & Industrial Engineering, 74(1), 111–119.

[47] Emily Ma, Yulin Liu, Jinghua Li, Su Chen. Anticipating Chinese tourists arrivals in Australia: A time series analysis. Tourism Management Perspectives 17 (2016) 50–58.

[48] Shaowen Li, Tao Chen, Lin Wang, Chenghan Ming. Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. Tourism Management 68 (2018) 116-126.

[49] Zhan R, Wan J. Iterated unscented Kalman filter for passive target tracking. IEEE Trans Aero Elec Sys 2007;43:1155-1163.

[50] Witt, S. F. and Witt, C. A., Forecasting tourism demand: A review of empirical research, International Journal of Forecasting, Vol. 11, 1995, pp. 447–475.

[51] Ma, H., and Zhang, Z., Grey prediction with Markov-Chain for Crude oil production and consumption in China, Advances in Intelligent and Soft Computing, Vol. 56, 2009, pp. 551–561.

[52] Vietnam National Administration of Tourism, Tourism Statistics, Retrieved on Mar. 10, 2012 from http://www.vietnamtourism. gov.vn/index.php?cat=2020

[53] Ouerfelli, C., Co-integration analysis of quarterly European tourism demand in Tunisia, Tourism Management, Vol. 29(1), 2008, pp. 127–137.

[54] Song, H. and Witt, S. F., General-to-specifi c modeling to international tourism demand forecasting, Journal of Travel Research, Vol. 42(1), 2003, pp. 65–74.FORECASTING MODELS IN TOURISM DEMAND 43

[55] Wang, Z., Liu, F., Wu, J., & Wang, J. (2014). A hybrid forecasting model based on bivariate division and a backpropagation artificial neural network optimized by chaos particle swarm optimization for day-ahead electricity price. Abstract and Applied Analysis, 2014, 31

[56] Sun JL, Zhang J, Ma HL, et al. Epidemiological features of typhoid/ paratyphoid fever in provinces with high incidence rate and in the whole country, in 2012. Zhonghua Liu Xing Bing Xue Za Zhi 2013;34:1183–8.

[57] Vollaard AM, Ali S, Van Asten HA, et al. Risk factors for typhoid and paratyphoid fever in Jakarta, Indonesia. JAMA 2004;291:2607–15.

[58] Bhan MK, Bahl R, Bhatnagar S. Typhoid and paratyphoid fever. Lancet 2005;366:749–62.

[59] Sun L, Shao Q, Wang ZQ, et al. Spatial structure of rodent populations and infection patterns of hantavirus in seven villages of Shandong Province from February 2006 to January 2007. Chin Med J (Engl) 2011;124:1639–46.

[60] Hsu, L. C. and Wang, C. H. The development and testing of a modifi ed Diff usion model for predicting tourism demand, International Journal of Management, Vol. 25(3), 2008a, pp. 439–445.

[61] Hsu, L. C. and Wang, C. H. Apply multivariate forecasting model to tourism industry, Tourism: An International Interdisciplinary Journal, Vol. 56(2), 2008b, pp. 159–172.

[62] Rob J Hyndman and George Athanasopoulos. Forecasting: Principles and Practice, Monash University, Australia

[63] J. faraway and C. chatfield.: Time series forecasting with neural networks: a comparative study using the airline data.[J], applied statistics, pp. 231–250, (1998).

[64] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov.: Dropout: A simple way to prevent neural networks from overfitting [J], Mach. Learn. Res., vol. 15, pp. 1929-195, (2014)

[65] Amigo JM, Hirata Y, Aihara K. On the limits of probabilistic forecasting in nonlinear time series analysis II: differential entropy. Chaos 2017;27:083125.

[66] Huang JC. Application of grey system theory in telecare. Comput Biol Med 2011;41:302–6.

[67] Crump JA, Luby SP, Mintz ED. The global burden of typhoid fever. Bull World Health Organ 2004;82:346–53.

[68] Lagos, D. G., Exploratory forecasting methodologies for tourism demand, Ekistics Journal, Vol. 396, 1999, pp. 143–154.

[69] Li, G. D., Wang, C. H., Masuda, S., and Nagai, M. A research on short term load forecasting problem applying improved grey dynamic model, Electrical power and Energy systems, Vol. 33, 2011, pp. 809–816.

[70] Pai P F, Hong W C.: An improved neural network model in forecasting arrivals [J]. Annals of Tourism Research, 32(4): 1138-1141(2005).

[71] Box, G.M. Jenkins.: Time series analysis forecasting and control [J]. Technometrics, 19:3, 343-344

[72] Yongkang Zheng.: The research for short-term load forecasting about Phase Space Reconstruction and Support Vector Machine [D]. Chengdu: Southwest Jiaotong University, 2008.

[73] Gan R, Chen X, Yan Y, et al. Application of a hybrid method combining grey model and back propagation artificial neural networks to forecast hepatitis B in china. Computer Math Methods Med 2015;2015:328273.

[74] Chau TT, Campbell JI, Galindo CM, et al. Antimicrobial drug resistance of Salmonella enterica serovar typhi in Asia and molecular mechanism of reduced susceptibility to the fluoroquinolones. Antimicrob Agents Chemother 2007;51:4315–23.

[75] Graves, A., Jaitly, N., Mohamed, A.: Hybrid Speech Recognition with Deep Bidirectional LSTM[J], Automatic Speech Recognition and Understanding (ASRU), (2013).

[76] Felix A. Gers, Jürgen Schmidhuber, Fred A.: Cummins Learning to forget: Continual prediction with LSTM Neural Computation, 12 (10), pp.

[77] Mahmod WE, Watanabe K. Modifiied Grey Model and its application to groundwater fflow analysis with limited hydrogeological data: a case study of the Nubian Sandstone, Kharga Oasis, Egypt. Environ Monit Assess 2014;186:1063–81.

[78] Bao CZ, Mayila M, Ye ZH, et al. Forecasting and analyzing the disease burden of aged population in china, based on the 2010 global burden of disease study. Int J Environ Res Public Health 2015;12:7172–84.

[79] Parry CM. The treatment of multidrug-resistant and nalidixic acid-resistant typhoid fever in Vietnam. Trans R Soc Trop Med Hyg 2004;98:413–22.

[80] Filippini, Massimo, and Lester C. Hunt. (2011) "Energy demand and energy efficiency in the OECD countries: a stochastic demand frontier approach." Energy Journal 32 (2): 59–80.

[81] Moutinho, L.: Consumer behavior in tourism[J]. Marketing, 21(10), 1–44(1987).

[82] Box G E P, Pierce D A.: Distribution of residual autocorrelations in autoregressive-integrated moving average time series models [J]. Journal of the American statistical Association, 65(332): 1509-1526(1970).

[83] M. S. Ahmed and A. R. Cook.: Analysis of freeway traffic time-series data by using Box-Jenkins techniques[J]. Transp. Res. Rec., no. 722, pp. 1-9, (1979).

[84] C. Lim and M. McAleer.: Time series forecasts of international travel demand for Australia[J]. Tourism Management,23, 389-396, (2002).

[85] Kenji Doya.: Bifurcations in the learning of recurrent neural networks[J].Proceedings of IEEE International Symposium on Circuits and Systems 1992, vol. 6, pp. 2777–2780(1992).

[86] J. Faraway and C. Chatfield.: Time series forecasting with neural networks: a comparative study using the airline data[J], Applied statistics, pp. 231–250, (1998).

[87] Herbert Jaeger.: Tutorial on training recurrent neural networks, covering BPTT, RTRL, EKF and the "echo state network" approach[J], German National Research Center for Information Technology, Technical Report GMD Report 159, (2002).

[88] Yoshua Bengio, Patrice Simard, Paolo Frasconi.: Learning long-term dependencies with gradient descent is difficult[J], IEEE Transactions on Neural Networks, 5 (2), pp. 157-166(1994). 2451-2471(2000).

[89] Tan, Y. F., McCahon, C., and Miller, J., Modelling tourist fl ows to Indonesia and Malaysia, Journal of Travel and Tourism Marketing, Vol. 12(1-2), 2002, pp. 63–84.

[90] Tsaur, R. C and Kuo, T. C, The adaptive fuzzy time series model with an application to Taiwan's tourism demand, Expert systems with Applications, Vol. 38, 2011, pp. 9164–9171.

[91] Wang, S. J., Wang, W. L, Huang, C. T., and Chen, S. C., Improving inventory eff ectiveness in RFID-enabled global supply chain with Grey forecasting model, Journal of Strategic Information Systems, Vol. 20, 2011, pp. 307–322.

[92] S. Hochreiter and J. Schmidhuber: Long short-term memory[J], Neural Compute., vol. 9, pp. 1735-17 0, (1997).

[93] Askari, M., and Fetanat, A., Long-term load forecasting in power system: Grey system prediction-based models, Journal of Applied Sciences, Vol. 11, 2011, pp. 3034–3038.

[94] Chang, Y. W. and Liao, M. Y., A seasonal ARIMA model of tourism forecasting: The case of Taiwan, Asia Pacifi c journal of Tourism research, Vol. 15(2), 2010, pp. 215–221.

[95] Kan, M. L., Lee, Y. B. and Chen, W. C., Apply grey prediction in the number of Tourist, The fourth international conference on Genetic and Evolutionary computing, 2010, pp. 481–484.42 T. L. NGUYEN, M. H. SHU, Y. F. HUANG AND B. M. HSU

[96] Huang, Y. L. and Lee, Y. H, Accurately forecasting model for the Stochastic Volatility data in tourism demand, Modern economy, Vol. 2, 2011, pp. 823–829.

[97] Jackman, M. and Lorde, T. Modeling and forecasting tourist fl ows to Barbados using Seasonal univariate time series models, Tourism and Hospitality Research, Vol. 10, 2010, pp. 1–13.

[98] Obaro SK, Iroh Tam PY, Mintz ED. The unrecognized burden of typhoid fever. Expert Rev Vaccines 2017;16:249–60.

[99] Liu FF, Zhao SL, Chen Q, et al. Surveillance data on typhoid fever and paratyphoid fever in 2015, China. Zhonghua Liu Xing Bing Xue Za Zhi 2017;38:754–8.

[100] O. Claveria, E. Monte, and S. Torra, "Tourism demand forecasting with neural network models: different ways of treating information," International Journal of Tourism Research, vol. 17, no. 5, pp. 492–500, 2015.

[101] S. Gao, Y Wang, Q Cao, Z Tang. Gravitational search algorithm combined with chaos for unconstrained numerical optimization, Applied Mathematics and Computation 231, 48-62, 2014

[102] T Zhou, S Gao, J Wang, C Chu, Y Todo, Z Tang. Financial time series prediction using a dendritic neuron model, Knowledge-Based Systems 105, 214-224, 2016

[103] S Gao, H Dai, G Yang, Z Tang. A novel clonal selection algorithm and its application to traveling salesman problem, IEICE transactions on fundamentals of electronics, communications and . . . , 2007

[104] S Gao, M Zhou, Y Wang, J Cheng, H Yachi, J Wang. Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction, IEEE transactions on neural networks and learning systems 30 (2), 601-614, 2019.

[105] L. Wang, Y. Zeng, and T. Chen, "Back propagation neural network with adaptive differential evolution algorithm for time series forecasting," Expert Systems with Applications, vol. 42, no. 2, pp. 855–863, 2015.

[106] K.-Y. Chen and C.-H. Wang, "Support vector regression with genetic algorithms in forecasting tourism demand," Tourism Management, vol. 28, no. 1, pp. 215– 226, 2007.

[107] M. Khashei, S. R. Hejazi, and M. Bijari, “A new hybrid artificial neural networks and fuzzy regression model for time series forecasting,” Fuzzy Sets and Systems, vol. 159, no. 7, pp. 769–786, 2008.

[108] O. Claveria, E. Monte, and S. Torra, “Tourism demand forecasting with neural network models: diffferent ways of treating information,” International Journal of Tourism Research, vol. 17, no. 5, pp. 492–500, 2015.

[109] L. Wang, Y. Zeng, and T. Chen, “Back propagation neural network with adaptive differential evolution algorithm for time series forecasting,” Expert Systems with Applications, vol. 42, no. 2, pp. 855–863, 2015.

[110] T Jiang, S Gao, D Wang, J Ji, Y Todo, Z Tang. A neuron model with synaptic nonlinearities in a dendritic tree for liver disorders. IEEJ Transactions on Electrical and Electronic Engineering 12 (1), 105-115, 2017

[111] Z Xu, Y Wang, S Li, Y Liu, Y Todo, S Gao. Immune algorithm combined with estimation of distribution for traveling salesman problem. IEEJ Transactions on Electrical and Electronic Engineering 11, S142-S154, 2016

[112] S. Gao, H Chai, B Chen, G Yang. Hybrid gravitational search and clonal selection algorithm for global optimization, International Conference in Swarm Intelligence, 1-10, 2013

[113] S. Gao, H Dai, J Zhang, Z Tang. An expanded lateral interactive clonal selection algorithm and its application. IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences, vol.E91-A, no.8, pp.2223-2231, August, 2008.

[114] S. Gao, Z Tang, H Dai, J Zhang. An improved clonal selection algorithm and its application to traveling salesman problems. IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences, vol.E90-A, no.12, pp.2930-2938, December 2007.