



中国海洋大学  
OCEAN UNIVERSITY OF CHINA

硕士学位论文



中国海洋大学  
OCEAN UNIVERSITY OF CHINA

# 硕士学位论文

MASTER DISSERTATION

论文题目: 青岛市房地产市场预测模型研究

英文题目: Study On Forecasting Model Of  
Real Estate Market In Qingdao

作者: 滕渊 (硕士)

指导教师: 刘强 副教授

学位类别: 全日制学术硕士

专业名称: 管理科学与工程

研究方向: 建筑工程管理



2013 年 05 月 20 日

谨以此论文献给我敬爱的导师刘强教授  
及关心和帮助过我的朋友们

-----滕渊



# 青岛市房地产市场预测模型研究

学位论文答辩日期：\_\_\_\_\_

指导教师签字：\_\_\_\_\_

答辩委员会成员签字：\_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_



## 独 创 声 明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的  
研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其  
他人已经发表或撰写过的研究成果，也不包含其他教育机构的学位或证书使用过  
的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的  
说明并表示谢意。

学位论文作者签名：                    签字日期：      年    月    日

-----

## 学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，并同意以下  
事项：

1、学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许  
论文被查阅和借阅。

2、学校可以将学位论文的全部或部分内容编入有关数据库进行检索，可以  
采用影印、缩印或扫描等复制手段保存、汇编学位论文。同时授权清华大学“中  
国学术期刊(光盘版)电子杂志社”用于出版和编入 CNKI《中国知识资源总库》，  
授权中国科学技术信息研究所将本学位论文收录到《中国学位论文全文数据库》。  
(保密的学位论文在解密后适用本授权书)

学位论文作者签名：

导师签字：

签字日期：      年    月    日

签字日期：      年    月    日



# 青岛市房地产市场预测模型研究

## 摘 要

中国的房地产业已经成为国民经济中的支柱产业，稳定健康的房地产市场对于国民经济稳定快速的发展有着不可或缺的作用。随着房地产市场不断的发展，很多问题也随之浮现。对于房地产市场目前存在的问题，各级政府在不断摸索中解决，在取得一定成效的同时也饱受各界人士的争议。很多房地产业内人士指出出现房地产宏观调控政策“头痛治头，脚痛治脚”的根本原因，是我国缺少合理的、完善的房地产市场预测和预警体系。

本文正是基于目前国内房地产市场现状，希望能建立一种行之有效的数学模型，对青岛市房地产市场加以模拟预测，并对其得到的数据加以分析和说明，为青岛市房地产市场预测模型的建立进行一次有意义尝试。

房地产市场与众多因素有着密不可分的联系，这也使得要建立房地产市场预测模型并不是简单。如果我们仅仅是使用单个预测模型来进行房地产市场相关指标的预测的话，往往不能取得良好的实际效果。所以本文准备运用两种不同的思想来建立模型，从趋势预测和因素分析两个方面来建立组合预测模型。本文首先运用改进型的灰色预测方法对相关房地产指标进行趋势性的预测，然后通过神经网络的思想建立模型从影响因素方面着手进行预测。最后运用信息论中的熵值法来组合这两个模型的结果。本文通过上述的方法建立的青岛市房地产市场组合预测模型比起常规的预测模型肯定具有一定优越性。以期用这种模型能够准确的预测青岛市房地产市场发展状况。

根据以上描述，在前人的研究基础上论文主要对以下几个方面的内容进行了探讨和研究：

（1）房地产市场预测指标如何选定：根据统计指标选定标准以及考虑到青岛房地产市场因素与数据的准确性和易得性，本文预测目标包括房屋施工面积、房地产开发投资额、房屋实际销售面积这三个指标。

（2）单项预测模型的选定：指标预测方法有很多种，哪几种预测方法对房地产市场预测比较适用？我们在了解各种预测模型思想的前提下，选取了灰色预测和神经网络预测方法，并在一些方面改进了这两种预测模型，使得改进后的预测模型能更好预测相关指标。



(3) 组合模型的选择：根据单项预测得到的预测结果，通过哪种方法组合起来比较合理？本文运用了信息论中的熵值法来判断各模型加权平均系数，使得该组合模型集合了灰色理论和神经网络的优点，模型反映出的所研究问题信息也更为全面。

**关键字：**房地产市场；灰色理论；神经网络；熵值法

# Study On Forecasting Model Of Real Estate Market In Qingdao

## Abstract

The real estate has become one of pillar industries of the national economy. Real estate market's stable healthy development plays an important role in the national economy, and let it has a fast and steady development. In recent years, since the development of the real estate market, many problems have aroused. For the present existed problem, all levels of government are finding solutions to these problems. These obtain the certain result, also appeared some deviations or the negative effect. Many experts point out that the root causes of macro-control of the property is not clear policy results in China takes on the situation of "taking stopgap measures" because of lack of a the reasonable and scientific forecast and early warning system of the real estate.

This article is precisely based on this background. We hope to establish an effective model to the predictions of real estate market in qingdao, analysis and explain the data we get. it is a meaningful attempt to build the prediction model of real estate market in qingdao.

We have observed a strong correlation between the real estate and many factors. so building the prediction model of real estate market will not be easy. In order to combine the advantages of various models and overcome their disadvantages, this paper proposes a combination model for forecasting based on conditional entropy. so the model was set up by employing the two kinds of thoughts. It can be viewed from two aspects: trend predictions and factor analysis.

The paper firstly constructs a model to carries on the trend estimate, which is used for an improvement gray forecast model; Then it constructs another model by neural network; Finally it set up a combination model for forecasting based on conditional entropy. The model show that the technical measures taken are feasible and with certain advantages. We hope the method can exactly predict the development of real estate market in qingdao.

Following the ideas above, This article in predecessor's research foundation, will carried from some ways:

- (1) how to specify predictors of real estate market: Judging from its selecting criteria and the accuracy and facility of data. the destination of predicting includes under construction floor space、 real estate development investment and commercial housing sales area.

- (2) how to selecte single prediction models: There are many kinds of prediction methods, which kinds of prediction method is applicable for real estate market forecasting? we select the grey forecasting and neural network forecasting to set up single prediction models, then construct the combining forecast model based on conditional entropy. With those modes, the system have very good characteristic.
- (3) how to selecte the combining forecast model: according to single prediction models,Which the combining forecast model is more appropriate. it set up a combination model for forecasting based on conditional entropy and the combination model set advantages of gray theory and neural network. the model is more suitable for the development of real estate market in qingdao.

**Keywords: real estate market; grey theory; neural network; entropy method;**

# 目 录

1 绪论.....	1
1.1 研究背景和意义.....	1
1.2 国内外研究现状.....	2
1.3 论文结构安排及研究方法和创新点.....	4
1.3.1 论文结构安排.....	4
1.3.2 研究方法和创新点.....	5
2 房地产市场概述及相关指标分析.....	7
2.1 房地产市场概述.....	7
2.1.1 房地产市场的含义.....	8
2.1.2 房地产市场类别.....	8
2.1.3 房地产市场的特征.....	9
2.1.4 房地产与国民经济的关系.....	12
2.2 青岛市房地产市场概况.....	13
2.2.1 青岛市概况.....	13
2.2.2 青岛市房地产市场概述.....	15
2.3 房地产市场预测的指标选定.....	19
2.4 本章小结.....	21
3 灰色预测模型.....	22
3.1 灰色系统与灰色预测概述.....	22
3.2 $GM(1,1)$ 的建模过程.....	24
3.2.1 原始数据预处理.....	24
3.2.2 $GM(1,1)$ 算法构造流程.....	25
3.2.3 $GM(1,1)$ 模型的特征.....	26
3.2.4 模型检验分析.....	27
3.3 $GM(1,1)$ 模型的局限性及改进.....	28
3.3.1 $GM(1,1)$ 模型的局限性.....	28
3.3.2 $GM(1,1)$ 模型的改进.....	29
3.4 改进 $GM(1,1)$ ——马尔柯夫模型.....	30
3.4.1 基于背景值和灰色作用量的 $GM(1,1)$ 模型改进.....	30
3.4.2 马尔柯夫模型.....	33

3.5 改进 $GM(1,1)$ ——马尔柯夫模型的应用 .....	34
3.5.1 改进 $GM(1,1)$ ——马尔柯夫模型建立 .....	35
3.5.2 实验结果分析 .....	39
3.6 本章小结 .....	40
<b>4 神经网络预测模型 .....</b>	<b>41</b>
4.1 神经网络简介 .....	41
4.2 人工神经网络模型原理 .....	42
4.2.1 人工神经元模型 .....	42
4.2.2 神经网络分类 .....	44
4.2.3 神经网络的学习规则 .....	45
4.3 BP 神经网络 .....	46
4.3.1 BP 神经网络结构及工作原理 .....	46
4.3.2 BP 神经网络学习算法的基本流程 .....	47
4.3.3 BP 神经网络的缺陷 .....	50
4.4 小波神经网络 .....	50
4.4.1 小波神经网络概述 .....	50
4.4.2 小波神经网络算法的基本流程 .....	52
4.5 BP 神经网络和小波神经网络模型的应用 .....	53
4.5.1 BP 神经网络的应用 .....	53
4.5.2 小波神经网络的应用 .....	55
4.6 本章小结 .....	57
<b>5 组合预测模型 .....</b>	<b>58</b>
5.1 组合预测的概念及意义 .....	58
5.2 熵值法组合预测模型 .....	59
5.3 基于组合预测的青岛市房地产市场预测 .....	61
5.3.1 组合模型的建立 .....	61
5.3.2 组合预测及结果分析 .....	63
5.4 本章小结 .....	66
<b>6 结论与展望 .....</b>	<b>67</b>
6.1 结论 .....	67
6.2 展望 .....	67
<b>参考文献 .....</b>	<b>69</b>

附 录.....	72
致 谢.....	77
个人简历.....	78
发表的学术论文.....	79



# 1 绪论

## 1.1 研究背景和意义

近些年房地产市场迅猛的发展为我国经济持续高速度的增长提供了强大的动力引擎，现在的房地产业不但是国民经济中的支柱产业，也是我国税收收入的重要来源渠道。而对于普通市民来说，住房可能是其一生最大数额的消费品，这需要他们十几年甚至几十年的收入的积累。一个良好发展水平的房地产市场对与国民经济稳定快速的发展有着不可或缺的作用。从行业影响系数来看，房地产业是一个影响系数很大产业，他带动了建筑、建材、银行、中介等一系列相关产业的快速发展<sup>[1]</sup>。

随着房地产市场不断的发展很多问题也随之浮现。房价上涨过快可能是普通人关心最多的问题了，目前北京、上海等一线城市的新房均价已经突破 20000 元每平方米，随便一套房可能就要上百万甚至上千万，很多家庭可能要花费一辈子的储蓄才能购得一套商品房。目前房地产市场投机行为十分严重，很多市民把房屋投资看作最佳投资渠道，以至于房价一直处于高位。国家也陆续出台了相关政策来抑制房价过快上涨的趋势，但都收效甚微。国家财政收入很大一部分是来自于地方土地买卖。土地是经过开发商竞拍的模式进入市场的。在竞拍过程中势必会抬高土地的价格，使得最终建造的房屋成本居高不下<sup>[2]</sup>。另外房地产市场秩序混乱，在这几年经济飞速发展的时期，国内半数以上的 100 强企业都进入了房地产领域，这一方面是因为房地产行业利润比较高，大家都想从中分一杯羹，另一方面也是因为目前的房地产市场制度不完善，准入门槛较低。总而言之，我国的房地产市场在高速发展的同时也存在了很多的问题。

国民经济和房地产业的关系可以用一句现在流行的话来描述：“房地产绑架了国民经济”<sup>[3]</sup>。现在的中国经济的高速发展已经离不开房地产了，用一句夸张的话说房地产就是国民经济的命脉。所以房地产业的兴衰牵动着各级政府的神经。对于房地产市场存在的问题，政府在不断摸索着解决，虽然取得了一定的效果，但也饱受各界人士的非议。很多房地产业内人士指出出现房地产宏观调控“头痛治头，脚痛治脚”的根本原因，是我国缺少合理的、完善的、及时更新的房地产市场预测和预警体系<sup>[4]</sup>。只有建立完善的房地产预警体系，才能精确的模拟和



预测未来的房地产市场发展动态，为政府出台准确合理的调控政策提供依据。房地产市场预警体系的核心环节就是市场预测模型，只有在建立准确有效的预测模型的基础上，才能对房地产市场相关指标进行预测，从而判断房地产市场的发展状况，为相关部门对管理房地产市场提供参考。

本文正是基于目前国内房地产市场发展现状，希望能建立一种行之有效的数学模型，对青岛市房地产市场加以模拟预测，并对其得到的数据加以分析和说明，为青岛市房地产市场预测模型的建立进行一次尝试。

房地产市场的发展与国民经济、自然环境、法律政策等很多因素有着密不可分的关系，这也使得要建立房地产市场预测模型并不能用很简单数学模型来描述。如果我们仅仅使用单个预测模型来进行房地产市场相关指标的预测的话，往往不能取得良好的效果。所以文本准备运用两种不同的方法来建立模型，从趋势预测和因素分析两个方面来进行组合建立模型。本文首先运用改进型的灰色预测方法对相关指标进行趋势性的预测，然后通过神经网络的思想建立模型从影响因素方法着手进行预测。最后运用信息论中的熵值法来组合这两个模型。本文通过上述的方法建立的青岛市房地产市场预测模型肯定具有一定优越性。以期用这种模型能够准确的预测青岛市房地产市场发展状况。

## 1.2 国内外研究现状

现代的房地产市场理论主要来源于发达国家，早期的房地产市场模型只包含简单的参数（如城市人口、工资收入等），这类模型对于现在房地产研究几乎已经没有了参考价值，但其理论分析仍然深深的影响着后人的研究方向。随着对房地产市场研究的深入，研究成果也越来越全面：E.W.Morehouse 和 R.T.Ely 在 1924 年出版的《土地经济学原理》一书中，阐述了土地与房地产发展的关系<sup>[5]</sup>。上世纪 60 年代 E.O.Olsen 开创了现代城市经济学流派，这个学派主要探讨了房地产市场和地理位置的关系<sup>[6]</sup>。20 世纪 80 年代，Clemen R.Tr 提出了从土地利用和土地供需方面来对房地产市场进行研究<sup>[7]</sup>。Mankiw N G 和 David N W 总结了房地产市场与经济周期的关系后，提出了房地产周期的分类，主要通过经济周期因素分析房地产发展状况<sup>[8]</sup>。上世纪九十年代对房地产市场的研究就出现了百家争鸣的情况：Box,G.E.P.主要通过泡沫理论进行研究房地产市场<sup>[9]</sup>；Anthony Y Gu.发现了房地产市场的发展与国民经济增长的内在关系<sup>[10]</sup>；Birch John W 主要通过市

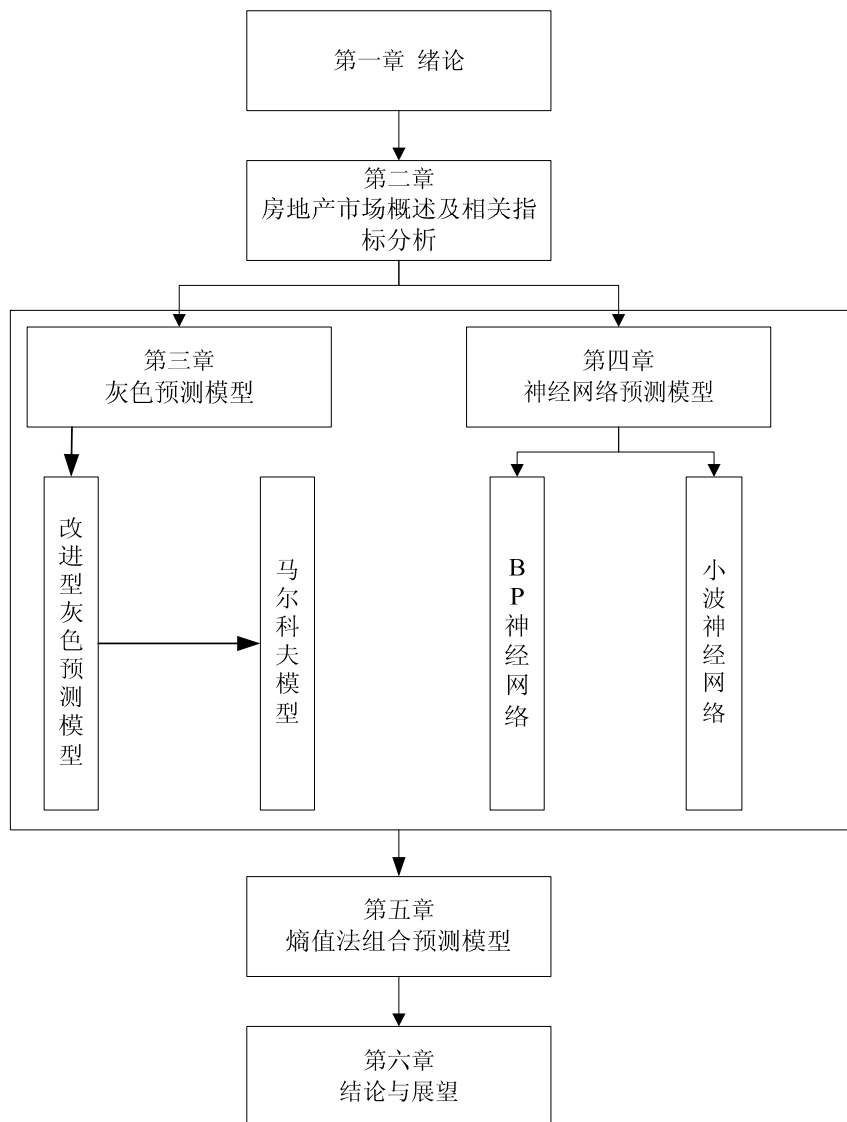
场运行理论来研究的<sup>[11]</sup>；S.w.Decaro 提出了通过房屋的供求关系、城市人口结构等因素来进行研究的<sup>[12]</sup>；Palmer PJ 主要通过空置率理论研究房地产市场发展状况的<sup>[13]</sup>。21 世纪以来相关的预测方法有了更为丰富的扩充，主要有以下几种：选取市场比较法、收益预测、信息统计等方法。

我国房地产市场预测模型的研究起步较晚，但随着近些年房地产市场飞速的发展，相关研究也逐渐的多了起来。它们主要通过以下几种理论进行研究预测的：

（1）逐步回归模型：这种模型基本思想是最小二乘法，只是通过分步骤实现的，通过加入删除变量的方法得到最终的指标。葛金辉、桑玉杰、孙斌艺等学者在这方面研究成果显著<sup>[14][15]</sup>。（2）神经网络模型：其模型基本思想是通过调整内部大量神经元之间的关系权值，使得不用求出具体的数学模型就能够通过输入相关初始数据，得到符合某种现象的预测值。它具有很高的自组织性和学习能力。李万庆、张金水、孟文清等学者在此方向有着深入的研究<sup>[16]-[19]</sup>。（3）灰色预测：其模型主要通过  $GM(1,1)$  模型对房地产相关数据进行趋势性预测。一般此类模型只是适用于历史数据波动较小的情况下。宋喜民，周书敬等学者发表了关于这一方面的文章<sup>[20]-[23]</sup>。（4）系统动力学模型：这种模型主要是解决非线性、多变量反馈的复杂模型，其实现手段主要是通过计算机仿真进行时变系统的研究。罗平、沈岐、胡雨村等学者进行了相关的研究<sup>[24]</sup>。（5）组合预测：其主要思想是通过某种数学方法求得各单项预测的平均权系数后组合形成一种新的模型。这种思想已经广泛的应用于各个领域，其中唐小我、王秋萍等学者就不同类别组合预测进行了研究<sup>[25]-[27]</sup>。

## 1.3 论文结构安排及研究方法和创新点

### 1.3.1 论文结构安排



本文内容共分为六章具体结构安排如下：

第一部分是绪论，对本文的研究背景、国内外研究现状进行了集中论述，论述了文章在现实应用中的意义，从而进一步的，提出了本文研究方法和结构安排。

第二部分是房地产市场概述和相关指标分析。这部分详细的阐述了房地产市场的基本概念、分类依据及固有特征，在此基础上总结了我国房地产业与国民经济之间的关系；在简单介绍了青岛市的概况及经济发展水平后，重点的介绍了近十年青岛房地产市场发展水平，最后通过按照相关指标选取的标准，选定了开发投资额、房屋施工面积、房屋实际销售面积三个指标作为预测目标。

第三部分是灰色模型预测。这部分首先介绍了灰色理论及其特征。给出了  $GM(1,1)$  建模思想和流程，指出了该模型的局限性和改进方式。最后给出了基于改进背景值和灰色作用量的新型  $GM(1,1)$  模型。通过此模型和马尔科夫模型的使用，预测了青岛房地产市场的相关指标，取得了良好的实际效果。

第四部分是神经网络预测模型。这部分通过对神经网络基础概念的介绍，对其结构原理、学习规则等网络基本要素进行了说明。在了解神经网络基本构造的基础上，介绍了 BP 神经网络和小波神经网络的工作原理和算法流程，简要的说明了两者的区别和联系。最后运用这两种神经网络分别对青岛市房地产市场进行相关预测。

第五部分是组合预测模型。这部分简单的介绍了两种组合预测的原理，给出了适用于青岛房地产市场组合预测的一种非最优组合预测方法——熵值法。通过运用熵值法把前两章使用的单项预测模型组合起来，通过三种模型比较发现该组合模型既有趋势预测的优点，又有因素预测的优势。运用该模型对青岛市房地产市场进行预测取得了较好的精度和应用价值。

第六部分是总结和展望。总结本文的研究结果，并对下一步工作进行展望。

### 1.3.2 研究方法和创新点

本文以青岛市房屋施工面积、开发投资额、房屋实际销售面积这三个指标为实例，先通过趋势分析，选定灰色预测的方法，通过改进背景值和灰色作用量构造了新模型，并结合马尔科夫模型对这三个指标进行预测研究，通过对目标值的影响因素分析，选取恰当的影响因素，结合神经网络与小波函数的优点，分别选取 BP 神经网络、小波神经网络对其历史样本数据进行训练并建立模型对这三个指标进行预测。为了更加全面的反映有研究对象的信息，以信息论中的熵值法为依据，计算出这三个模型的加权平均系数，建立关于青岛市房地产市场非线性组合预测

本文的创新点主要有以下几个方面：

(1) 本文没有用住房供应量，住房需求量，房价等指标描述房地产市场的发展状况。而是用更为基础的，没有任何争议的指标房屋施工面积、开发投资额、房屋实际销售面积这三个指标来进行预测的。其实这三个指标也分别与房屋供应

量、经济景气指数、房屋需求量有着莫大的关系。这三个指标的选定对房地产市场的预测模型的现实意义不言而喻。

（2）在灰色模型的建立过程中，发现了很多不合理的地方。文章从三个方面来改进了传统的灰色预测模型：初值，背景值，灰色作用量。并且结合使用了马尔科夫模型。这使得模型在建模依据上更为合理。

（3）本文运用了趋势预测（灰色预测和 **BP** 神经网络）和因素预测（小波神经网络）的组合预测模型，通过熵值法把两种不同性质的预测方法的优点有机的结合了起来。

## 2 房地产市场概述及相关指标分析

### 2.1 房地产市场概述

房地产是经济学中一个常见的概念,要研究相关房地产的发展状况必须要求我们对这一概念有着深入的了解。在中国人们一般不区分“房地产”和“不动产”的含义,而在港台地区一般更倾向于使用“不动产”这个名称。在台湾地区政府统计报告也把树木归于“不动产”中,香港地区通常把房产和土地统称为“地产”。从表面上来看房地产就是由房产、地产以及房产和地产所带来的各种权益形成的一个组合词。就具体情况来说,房产是指在土地上的各种建筑物,其中包括住宅、公寓、写字楼、厂房等,地产是指地上和地下范围内,各种建筑设施。它从存在的形式上可分为三类:土地(土地使用权)、建筑物和房地结合的房地产。因此我们就可以这么理解房地产这个概念:土地和土地上永久性建筑物及其衍生的权利和义务的总和。

从实物形态上来说,房地产是房产和地产的统一体。平时我们所说的土地概念广泛,只有当土地作为建筑用地及配备相关设施使用的时候,才构成房地产的组成部分。现实生活中的房屋不可能凭空而建,它们必须是依附在相关土地上的,所以说实物形态的房产是离不开土地的,我们必须把房产和地产作为一个与土地的相关的概念来看待。从价值形态上来说,房地产是一种特殊的商品,它肯定是在存在一定的使用价值和价值。在商品流通过程中,它消耗着人类的一般劳动,反过来也满足了人们的特定的需求,这就是房地产的价值和使用价值的所在。我们可以把整个房地产项目的经营过程看作社会再生产的过程,房屋经过建设、销售和使用直至最终的报废,然后再经过建设的过程就可以看作房地产市场的一个完整的循环过程。在整个循环过程中必定构成了与房地产业相关的经济关系。从产权关系上来说,房地产不仅是社会财富的代表,也反映了社会成员的经济权益关系。因为房地产和土地密切相关,而土地在现实生活中是不能够移动的,所以房地产在交易过程中只是把房地产的所有权、使用权等权益进行了转移。同理房屋的租赁只是把房屋的相关使用权进行了调整。正是因为房地产的相关权益如此复杂,所以它与一般的商品交易有着很大的不同,为此国家出台了很多关于房地产

交易的法律法规，规范房屋交易过程，保证房地产市场的繁荣和稳定。总的说来，在经济学上来说房地产既是相关实物的结合体，也是商品属性的聚合体，而且体现着社会成员的经济关系及权益<sup>[28]</sup>。

### 2.1.1 房地产市场的含义

从定义来说，市场有着很多种说法例如：“市场是商品交易的场所”、“市场是商品交易的过程”、“市场是需要某种商品或者服务的目标客户”。在我看来，市场是商品交易和社会发展的产物，它既包括买卖双方，也包括双方的交易过程和交易场所。所以综合来看，市场就是商品交换活动的构成和交换关系的总和。

房地产市场就是房地产商品交换活动和交换关系的总和，它主要有三种具体方式：房地产买卖、房地产抵押、房地产租赁。其中房地产买卖是指相关房屋的使用权和所有权的变更。房地产抵押是指单位和个人为了取得一定数额的贷款而出让房地产作为如期还款的抵押保证物的活动行为。房地产租赁是指房地产所有者在一定的期限内通过向承租人转让房地产相关的使用权益，而取得收益的活动行为<sup>[29]</sup>。当然房地产市场和普通意义上的市场一样，不光有房屋买卖双方，也包括双方买卖的交易过程和交易场所。这样看来，房地产市场是指整个房地产交易关系的总和，它不仅包括参与房地产交易活动人员、实物和活动本身，也包括在此过程中的相关信息流通，技术指导、资金流转等一系列的交易服务性的活动。

### 2.1.2 房地产市场类别

从房地产的建设过程层面上划分，房地产市场可以划分为一级、二级、三级三个既相互联系又相互区别的概念。房地产的一级市场也叫做土地市场，是指国家按照相关法律法规，通过政府相关部门把国有土地或者集体土地出让给房地产市场的行为。这些出让的土地可以是直接达到开发条件的土地，也可以是需要经过改造才能使用的土地。在我国土地所有权属于国家，个人只有使用权，所以土地市场是由国家垄断控制的。并且在土地市场上获得的土地有一定的时间期限。房地产的二级市场是指开发商通过对所取得的土地进行相关的建设，形成房地产实物状态，转让房地产相关权益的过程。一般来说是指商品房进行首次交易从而形成的房地产二级市场。当然这个过程也包括土地的二级市场，即土地拥有者按

照相关法规再次进行土地转让,使未开发的土地再次进入市场流通领域。由此看来房地产二级市场的交易双方是开发商和普通消费者,商品房的出售、先租后售以及预售过程都是属于二级市场的一部分。房地产三级市场是指普通的房屋拥有者和单位在流通市场中出让其拥有权或者使用权的过程。通俗一点来说房地产的三级市场也就是我们所说的二手房市场<sup>[30]</sup>。由此看来房地产三级市场的交易双方是在消费者与消费者,二手房的出售、抵押、出租等行为都是属于三级市场的。

在房地产市场中,一级市场、二级市场与三级市场是相互关联,相互渗透有着很强的联动性,从而组成了一个完整的房地产市场生态圈。例如:一级市场的土地价格肯定会传递到二级市场的商品房价格,从而影响三级市场的二手房商品价格,反之三级市场的二手房和二级市场的商品房交易状况也直接影响一级土地市场的出让行情。

### 2.1.3 房地产市场的特征<sup>[31]</sup>

由于房屋本身具有固定性、高价值性等特有属性,这使得房地产市场有着不同与其他商品市场的特征。如果要进行相关房地产市场的预测就必须首先深刻了解房地产市场的基本特征,这样才能的对房地产市场进行准确的把握,从而使得开发商可以对房地产开发规模和投资幅度进行合理的调整,也为政府进行相关市场调控提供的实实在在的依据。

#### (1)位置的固定性

房地产的位置固定性的特点是因为房屋建筑土地是不可移动的及土地上的建筑物也是固定的。对于一般商品而言,地域的特性并不直接构成商品的价值和使用价值,而对于房地产市场来说,同一个地区不同的地段即使相同品质的房屋的售价也是天差地别的。另外由于地域的固定性也决定了房地产市场不是自由流通的市场。当一个地区的房地产市场供求失衡或者价格差异很大的时候,并不能让房屋通过在不同地域的流转使得这种不平衡或者差异消失。房地产的这个特性使得不同地区的房地产市场有着很大的差别。由于不同地区的经济,人口、环境等因素的不同而使得不同城市的房地产市场有着不同的发展水平。一般来说经济发达的城市比经济落后城市的房地产市场活跃。甚至同一个城市不同地块也受各种因素导致其发展水平差异很大,另外文化、政策、城市规划等因素也会影响房



地产市场发展状况。比如：长白山、海南的旅游房地产市场，义乌、晋江的小商品房地产市场等。

## (2)发展的周期性

房地产业在国民经济中占有很大的比重，类似于经济发展具有很强的周期性，房地产市场也有着明显的周期性。它和宏观经济发展趋势密切相关，也与宏观经济的与该行业的协调程度紧密关联，这使得房地产市场随着经济发展周期波动而波动。一般来说，经济不景气的时候，投资者和消费者都会对房地产市场持有观望态度，投资考虑会更加慎重，这可能导致与房地产出现萧条不景气的现象；在经济形势发展良好的情况下，投资者就会扩大投资规模，对于投资项目就会比较乐观，消费者也乐于参与房地产交易。这可能会使得市场比较活跃。一般我们可以描述房地产的周期性为：繁荣阶段—衰退阶段—萧条阶段—复苏与增长阶段—繁荣阶段。

从我国房地产发展的实际情况来看，我国的房地产市场也具有明显的周期性，大体的时间与国民经济发展周期基本吻合。1981-1986年市场上的房屋供应量小于需求量，政府以及相关投资者不断向房地产市场投入资金；1987-1990年市场明显供过于求，需求不断下降，房屋空置率也在不断上升；1990-1997年市场处于悲观状态，消费者和开发商都对市场抱有过于谨慎的态度，1998-至今，随着国民经济不断的发展我国房地产市场迎来了发展新的高峰，至今仍然处于平稳上升的阶段。总的看来房地产的周期性表现为其增长、繁荣、衰退、复苏、理性发展等不同阶段的发展特征，这对于政府、开发商和普通市民都有着很重要的指示性的作用。对于政府来说必须充分深入掌握房地产发展的周期规律，建立相应的指标体系和预警系统，对其发展实时监控，及时出台合理的调控政策，使得房地产市场能够在健康、有序的轨道上发展。

## (3)政策的敏感性

房地产业不但是国民经济中的支柱产业，也是我国税收收入的重要来源渠道。对于普通市民来说，住房可能是其一生最大数额的消费品，这需要十几年甚至几十年的收入的积累。所以说不管什么时期，我国政府都会对房地产业进行强有力的政策干预，这也往往使得房地产市场对政府出台的政策比较敏感，近些年我国的房地产市场表现的尤为严重，比如：发展房地产需要大额的资金投入，这

使得房地产业不可能离不开银行信贷的支持，政府往往一进行相关的信贷政策的调整，房地产市场就反映相当剧烈。无论是开发商的项目开发资金，还是普通市民的按揭条款都被国家出台的信贷政策深深的影响着。可以这么说我国目前的房地产市场制度还不是很成熟，与发达国家的房地产市场相比，我国的房地产市场就是一个“政策市场”。政府出台的政策一般是通过以下几个方面对房地产市场进行影响的：土地供应和城市规划、抵押贷款政策、户籍政策、税费政策等。

#### (4)不完全竞争性

目前我国的房地产市场不是完全的竞争市场，一般来说一个完全竞争的市场具有以下特征：商品资源完全自由流通、市场信息透明、商品是同质的、有足够多的参与者。因为房地产位置的固定性决定了房屋是不能够完全自由流通的；在房地产交易过程中有很多政策法规的制约，市场信息不能够也不可能完全公开，这导致买卖双方所掌握的信息严重不对称；当然地理位置、建筑成本、规划设计等相关因素不可能使得建造出来的房屋是同质的；另外我国的房地产市场是有限开发的市场，个人没有土地开发权，商品房只能有相关具有资质的企业进行开发销售。房地产市场的这些固有特征使得它与完全竞争市场还是有很大的差别的，所以我国的房地产市场不是也不可能是完全竞争的市场。

#### (5)资金依赖性

房屋本身的价值和使用价值决定了房地产业的发展需要巨大的资金推动，单从一个房地产开发项目上来说，一般需要几亿、十几亿甚至更多的资金投入，如果完全靠房地产公司本身来承担这些资金是不可能的。这就需要银行、相关基金、金融公司等一系列金融机构的支持。发达国家的房地产市场融资比例在 35%左右，而我国房地产市场资金来源过度依赖银行，超过 65%的资金来源于银行的抵押和贷款。08 年的美国次贷危机归根结底是因为房地产市场对银行业资金的过度依赖，金融市场的波动对房地产市场有着决定性的影响。所以只有规范金融秩序，完善金融制度，才能使得房地产业在正常的轨道上发展。

#### (6)市场投机性

一般来说，市场的存在就决定了投机的存在。在我国房地产市场很大一部分投机行为是由于对于土地的投机造成的，因为土地资源是稀缺的，可供房地产开发的土地就更少了。这使得房屋的价值随着土地的价值存在而一直存在，具有很

高的保值增值效益，所以这也导致目前的房地产市场中投机性很强。如果能够深入研究房地产市场发展规律，在房地产市场中进行投机行为，这就会对房地产市场造成很严重的冲击。以前我国房地产相关制度法规不完善，导致投机行为到处可见，近些年国家加快了相关法律法规的制定，出台了一系列的抑制投机行为的政策法规，使得房地产业朝着健康、稳定的方向迈进。

#### 2.1.4 房地产与国民经济的关系

目前我国房地产在国民经济中占有很大比重，是国民经济的重要组成部分。在社会经济发展过程中，国民经济与房地产即相互制约又相关促进，良好的房地产市场发展水平对与国民经济稳定快速的发展有着不可或缺的作用<sup>[32]</sup>。从行业影响系数来看，房地产业是一个影响系数很大产业。它的发展很大程度上决定了其他行业的发展趋势，进而促进了整个社会的经济发展水平。房地产商品在开发过程中需要很多物质材料和技术支持，与相当一部分行业相关。房地产业的发展能够为相关产业在扩大规模、增加收入、完善产业链的进程中提供极大的帮助。总的说来，房地产在国民经济中处于支柱、先导的地位决定它在推进城市化进程、促进社会经济增长等方面有着极大的作用。从国民经济的角度来看，房地产业与国民经济不是简单的局部与整体的关系，它们之间有着很多内在的联系。国民经济的发展状况是房地产行业的发展的前提和基础，经济大环境制约或者促进着房地产市场的发展进程。此外房地产业也是国民经济的重要组成部分，它的发展状况也直接关系着国民经济的发展是不是平衡和结构是不是合理。所以房地产能否稳定健康的发展的关键在于能否正确协调处理两者之间的关系，同时其发展状况也影响着国民经济能否持续的增长、产业结构能否合理的调整等重大问题的处理。

第一、房地产是国民经济发展的基础保证和物质前提。房地产为全社会所有部门提供了物质空间，更是各个生产部门不可缺少的基本生产要素。社会的进步以及现代城市化进程的不断推进都是房地产业为基础的，这些都体现了房地产的基础性地位。

第二、房地产业对其他众多产业有着直接或者间接的影响。在我国房地产投入 1 元就能在其他 26 个相关产业产出 1.56 元，其中与之间接相关的产业多达 100

多个。例如我国房地产业 2005 年消耗钢材是全国产钢量的 14.67%，消耗玻璃占比多达 41.25%，消耗水泥占比 46.89%，运输总量占比 9.8%。以 2005 年青岛商品房销售额为 267 亿元为基数所带动的青岛全社会商品销售额为 356 亿元，占当年青岛市全社会消费总额的 26.24%<sup>[33]</sup>，因此可以看出房地产业引导和影响着众多相关行业的发展。

经过 30 多年的改革开放，我国国民经济得到了长足的发展，其中房地产作出了相当大的贡献。一个健康稳定的房地产市场可以持续的拉动建材、家电、装饰等数百个行业的不断发展，进而影响国民经济的可持续发展。可以看出不管是微观还是宏观上来讲，我国的房地产业和国民经济之间协调的发展是推进中国现代化社会主义进程不可或缺的环节。如何处理房地产在国民经济中的定位，平衡与其他产业的地位、以及两者之间过度依赖的问题都考验着相关部门监管和调控水平，为此完善房地产市场机制、加强合理的政策调控是处理两者之间的问题的必不可少的手段。所以说要实现国民经济和中国房地产业和谐的发展必须通过一系列的政策创新，一方面完善国民经济结构，另一方面规范和引导房地产业的发展。只有这样才能使得国民经济和房地产业得以长期稳定的发展。

## 2.2 青岛市房地产市场概况

### 2.2.1 青岛市概况

青岛市位于山东胶州半岛南端，西北连接陆地，东南濒临黄海，是山东省主要的区域经济中心，东北亚航运中心，国际滨海旅游度假胜地。青岛市现辖七个区(市南、市北、四方、李沧、崂山、城阳、黄岛)，五个县级市(平度、莱西、即墨、胶州、胶南)，总面积 10674 平方公里，总人口 725.65 万人，青岛海岸线长 133 公里，沿海的滩涂面积 376.4 平方公里，现有海岛 69 个，沿海旅游资源丰富。青岛是目前中国举办大型赛事和国际盛会最多的城市之一<sup>[34]</sup>。在举办了 2008 年奥帆赛之后，作为 2008 年北京奥运会唯一合作伙伴城市的青岛迎来了快速的发展阶段，在经济、文化等各方面取得了喜人的成果。

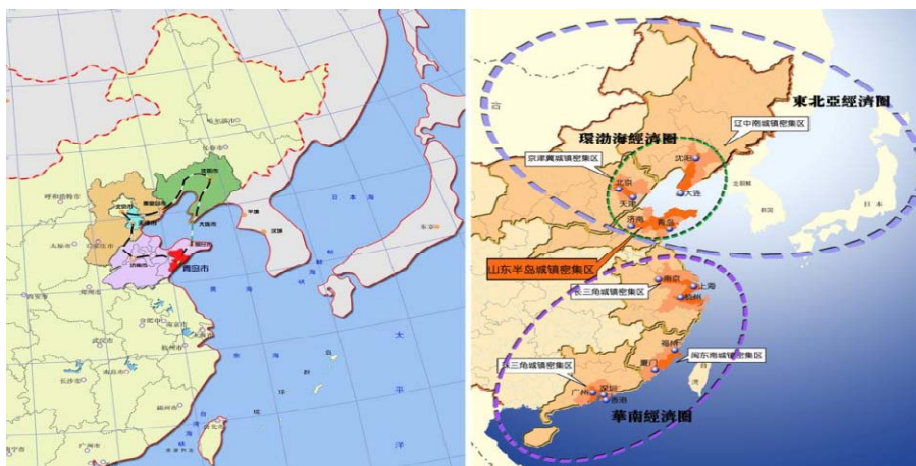


图 2-1 青岛市地理位置

青岛是山东的东部沿海现代化工业城市，在城市发展过程中不断通过技术升级、引入高尖端技术等核心战略，使得青岛成为山东乃至全国很重要的经济增长区域。目前环渤海湾经济圈形成了以青岛为主要出海口的胶州半岛地区的城市体系，对内以华东平原为腹地，对外以日韩为主要联系方向。

青岛在华东地区发展过程中担当了排头兵作用，工业上聚集了核心技术发展了高新产业，同时分散着传统和初级产业链，就全国范围来看青岛的城市经济辐射能力很强，工业化水平处于先进行列。近些年青岛充分发挥了其海洋产业化中心地位，在工业上形成了以现代制造业、高新技术产业为核心的发展模式，在经济上成为了区域贸易、信息中心，其城市化进程大大加速，成为全国著名的海滨旅游城市。近些年青岛的经济增长速度大幅以领先于全国平均水平。根据相关数据显示，2010 年 1-6 月份青岛市实现全市生产总值 2787 亿元，同比增长 14%，增幅较去年同期加大 3.5 个百分点。2000 年以来青岛市平均经济增长速度 14.7%，较全国水平高 4.8 个百分点。2010 年青岛市城镇居民收入为 24367 元，超出同期全国水平的 1.29 倍，人均 GDP 为 67059 元是同期全国水平的 2.57 倍，基本达到中等发达国家水平。但是与沿海发达城市相比还是处于中下游位置。从城市发展进程来看青岛具有很大的潜力。青岛市人口比较稳定，2000-2010 年间全市户籍人数基本保持在年均 3‰的增幅，其中 2010 年中心城市人口为 276.85 万人，占全市全部人口的比重为 36.4%。较 2001 年提高了 2.7 个百分点。目前青岛经济以第二产业为主，但第三产业发展迅猛。就 2010 年来看青岛第一、二、三产业所占比重分别为 4.1%，49.2%和 46.7%。第一、二产业增长率略显下降趋势，第三

产业同比增长明显，2010 年青岛第一产业同比增长 3%，第二产业增长 12.7%，第三产业增长 16.8%。

青岛市的经济一直走在全国发展的前列。全市的生产总值占全省生产总值的比重在不断上升，产业结构也日趋合理。但是仍然存在一些问题值得我们注意：总体上看，工业的整体素质和水平不高，传统产业所占比例较大，而丰富的海洋资源、科技优势还未能完全地转化为经济优势和竞争优势；相关的政策扶持力度有待于进一步加强；现行的城市管理理念不能适应新形势的需要，同时基础设施建设相对薄弱，城市的防灾减灾能力、污染防治等问题突出。

### 2.2.2 青岛市房地产市场概述

近年来，随着青岛城市化进程的不断推进，青岛市房地产市场以稳健的脚步向前迈进，目前也逐渐成为了全国房地产 40 个热点城市之一。进入新世纪以来，青岛市经济综合实力和城市核心竞争力不断加强，使得青岛房地产市场进入了新一轮高速发展期。下面我就简单的介绍以下 2000—2010 年十年内青岛房地产市场的具体情况<sup>[35][36]</sup>。

2009 年，青岛全市（包括市内 4 区，市郊 3 区）土地购置面积为 301 万平方米，开发面积为 379 万平方米，较上年分别下降 15.4%和 17.2%。由图 2-2 可以看出，2000-2009 年全市土地购置面积和开发面积的增长率波动较大。其中 2006 年是青岛市土地开发面积和土地购置面积最多的年份，分别为 948 万平方米和 749 万平方米。

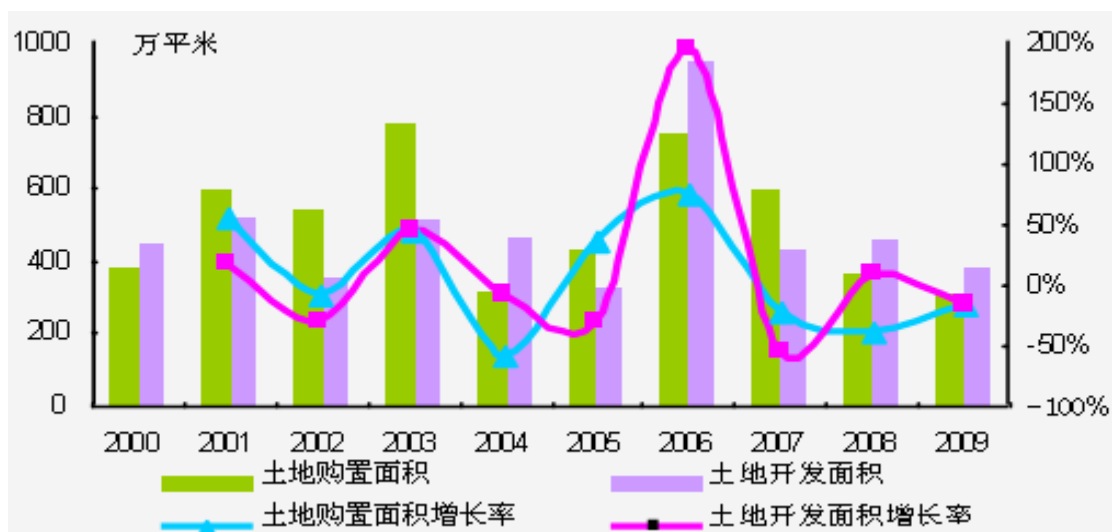


图 2-2 2000—2009 年青岛市土地购置与开发面积及其增幅

2001-2009 青岛对其房地产开发投资额年均增长 23.8%，比全国同期平均水平低了近 0.6 个百分点。仅在 2005、2009 年这两年中其平均增长率超过了全国同期平均水平。2009 年青岛全市（包括市内 4 区，市郊 3 区）房地产开发投资额为 459.4 亿元，同比增长 23.1%，较全国水平相比高出了 4.4 个百分点。

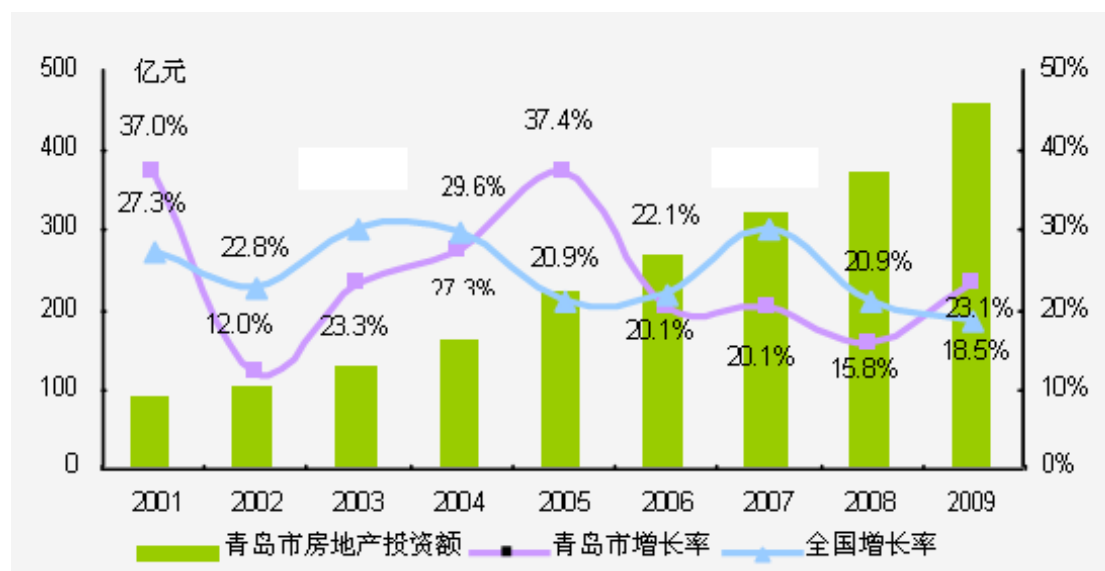


图 2-3 2001 年—2009 年青岛市房地产开发投资额及其增长率

2001 年—2008 年青岛市房地产市场开发投资占国内生产总值的比重略低于全国平均水平。2005 年以来，青岛市房地产市场开发投资占国内生产总值的比重一直稳定在 8.4% 附近，但是同期全国房地产投资占国内生产总值的比重已由 2005 年的 8.6% 不断上涨至 10.2%。从图 2-4 可以看出 2001-2008 年间青岛市房地产开发投资比全国平均水平大概低了 2 个百分点左右。

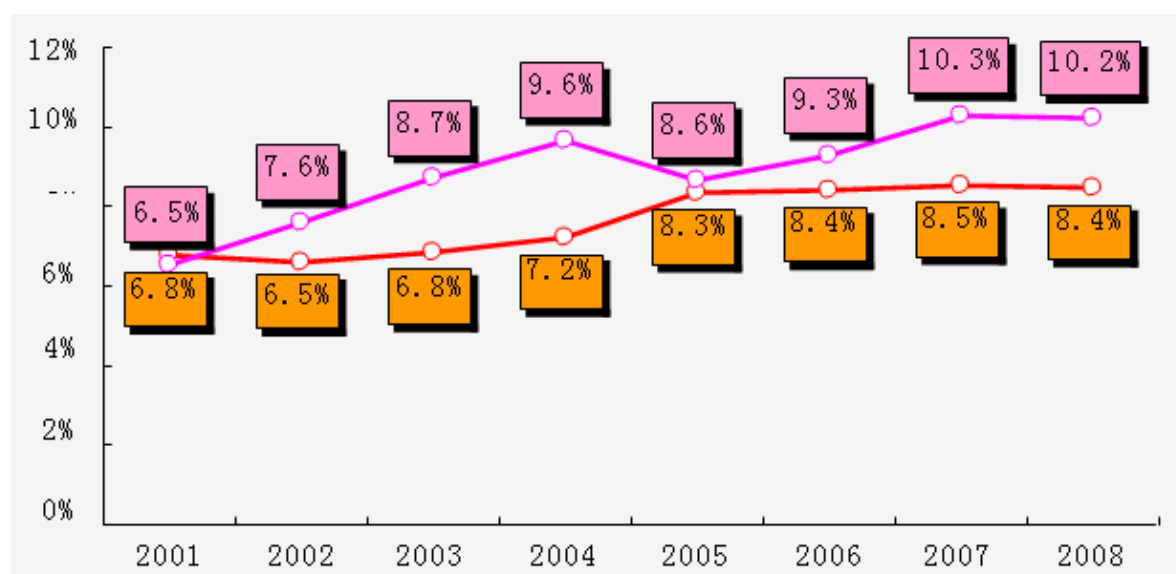


图 2-4 2001 年—2008 年房地产市场开发投资占国内生产总值的比重

从图 2-5 可以看出 2003-2009 连接青岛市房屋施工面积增长率平稳, 维持在 17% 左右, 但房屋新开工面积增长率波动较大, 平均年增长率为 10% 左右。由于金融危机的影响 2008 年全市房屋新开工面积增长率首次出现了负增长。2009 年青岛市房屋施工面积为 4310 万平方米和房屋新开工面积为 1331 万平方米, 较去年全市同期分别增长了 16.3% 和 16.2%。

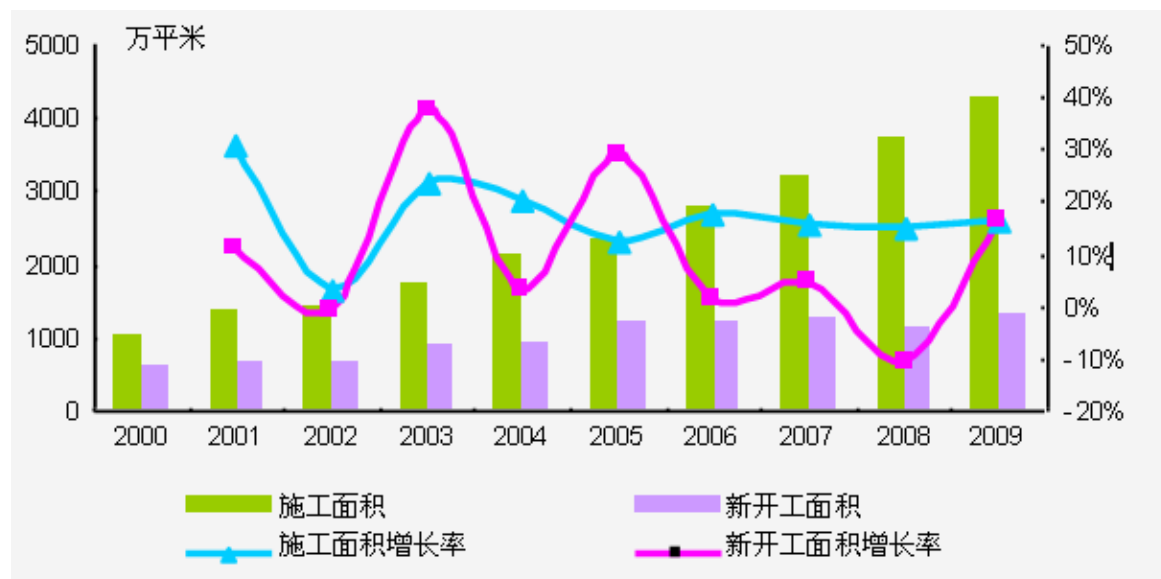


图 2-5 青岛市商品房施工和新开工面积及其增长率

2009 年青岛全市（包括市内 4 区，市郊 3 区）房销售面积为 1261.87 万平方米、竣工面积分别为 814.25 万平方米，销竣比明显大于 1。2009 年青岛商品销售面积与竣工面积之比自 2006 年开始连续四年大于 1，其中 2009 年比值达历史最大值 1.55。

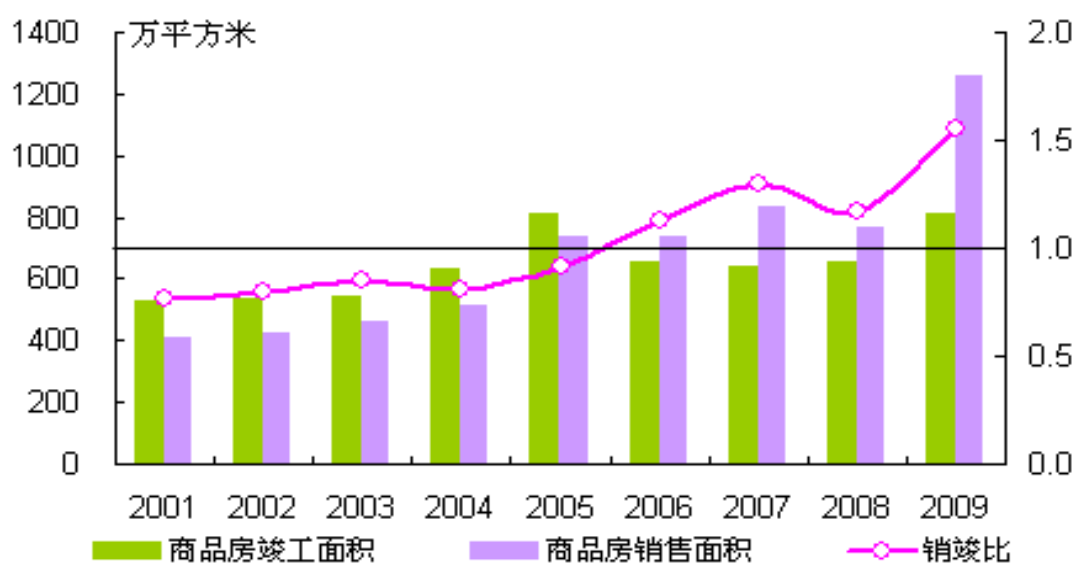


图 2-6 2000—2009 青岛市商品房销售与竣工面积



从图 2-7 可以看出自 2000-2009 年商品房销售面积和销售额在稳步增长。其中 2009 年商品房销售面积为 1262 万平方米，销售额为 704 亿元，较去年全市同期分别增长 63% 和 81%。从图 2-7 可以看出从 2005-2009 青岛市商品房销售面积和销售额增长率波动比较大。

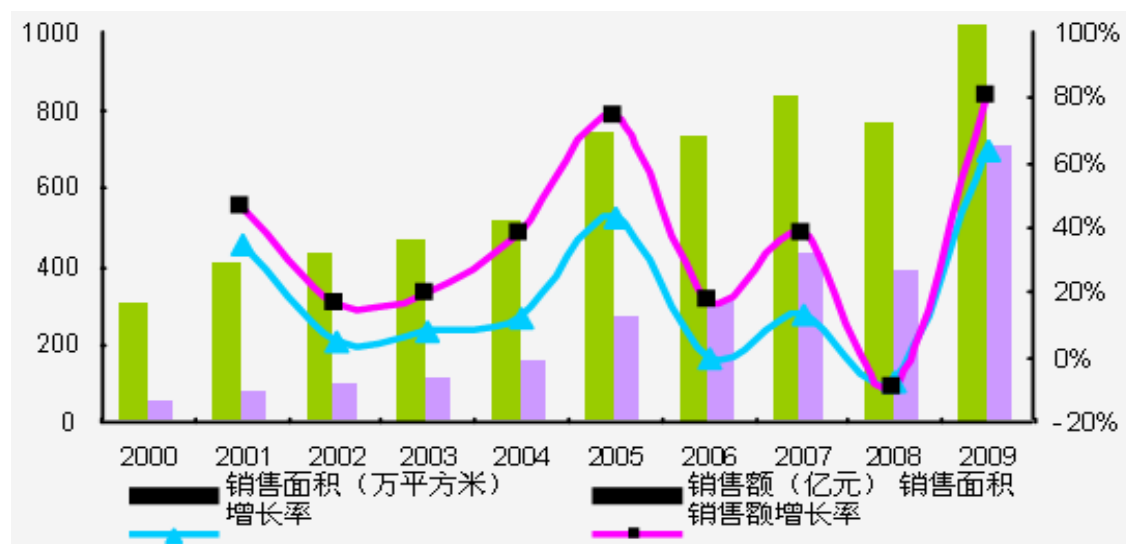


图 2-7 2000—2009 年青岛市商品房销售面积和销售额及其增长率

2000-2007 年青岛全市（包括市内 4 区，市郊 3 区）房屋销售均价增长率一直高于全国同期平均增长水平，但 2009 年青岛市房屋销售均价同比增长 9.9%，与全国同期增长率相比低了 10.8 个百分点。

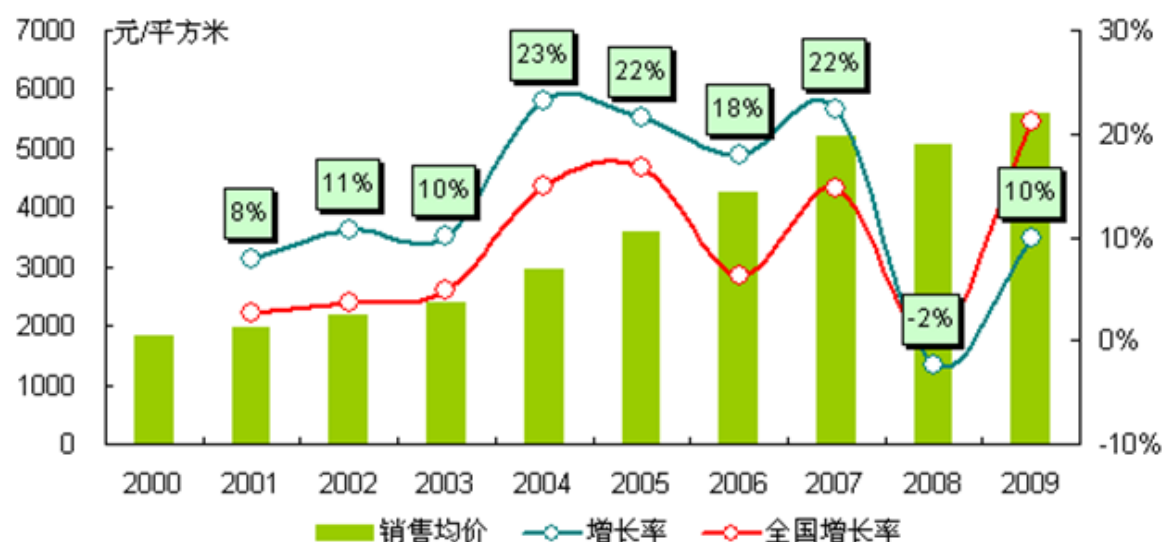


图 2-8 2000—2009 年青岛市商品房销售均价及其增长率

2001-2008 年期间，青岛市城市居民人均住房建筑面积不断增大。2008 年，青岛市城市居民人均房屋居住面积为 26.6 平方米，与全国 40 个重点城市相比，人均居住面积大于北京、上海、深圳等国内一线城市，但相比济南、郑州等二线

城市人均居住面积偏小。

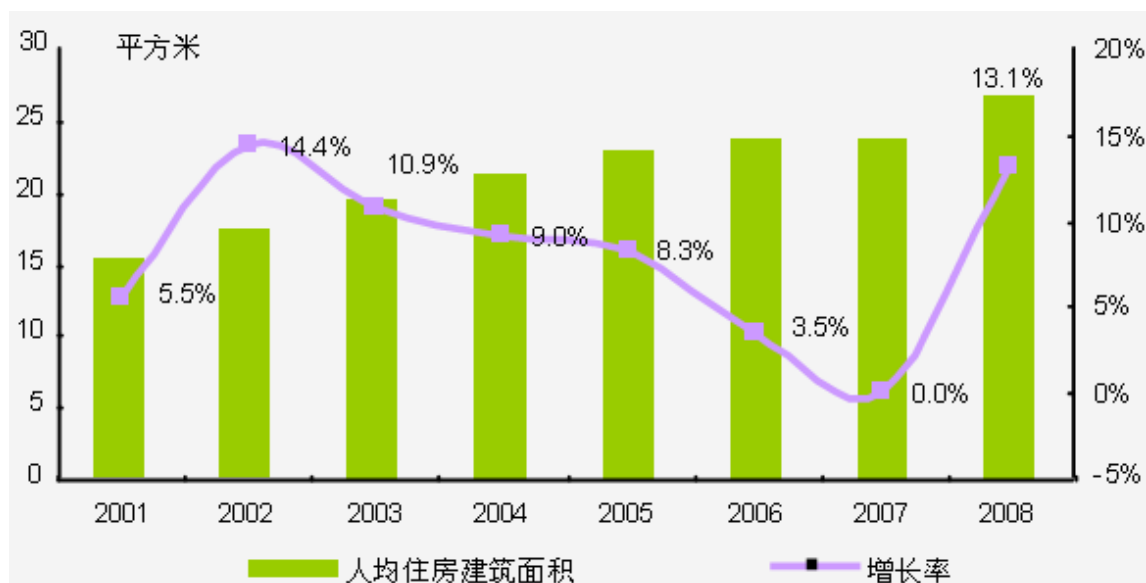


图 2-9 2001—2008 年青岛市人均住房建筑面积及其增长率

## 2.3 房地产市场预测的指标选定

近些年随着我国国民经济的不断发展,我国的房地产产业得到了迅猛的发展为我国经济持续高速的增长提供了强大的动力引擎,于此同时也带动了建筑、建材、银行、中介等一系列相关产业的快速发展。在享受房地产业带来巨大利益的同时一些问题就随之浮现出来,比如:房价上涨太快、住房供应模式单一、房地产市场秩序混乱、银行信贷政策不稳定等。为了解决这些突出性的问题,各级政府不断出台了关于税收、信贷、保障房等一系列的宏观调控政策,积极规范和疏导房地产市场秩序。虽然这些年出台的相关房地产调控政策取得了一定的效果,但也饱受各界人士的批评。很多房地产业内人士指出出现房地产宏观“调控头痛治头,脚痛治脚”的根本原因,是我国缺少合理的、完善的、及时更新的房地产市场预测和预警体系。没有完整的房地产市场统计数据就不能对房地产发展状况进行预测判断,也就无法形成完善的房地产市场预警体系,更加不可能出台恰到好处的楼市宏观调控政策,由此看来房地产市场统计数据就显得尤为重要。在这里我们首先要介绍以下统计指标的相关概念。一些房地产统计指标虽然不能完全反映房地产市场真正的发展趋势,但是确实能给房地产市场状况进行一定的说明,接下来我们首先了解一下相关统计指标状况。

首先市场统计指标往往是反映经济现象在数值方面的特性。一般来说一个社会经济现象表述的规范性需要统计指标来说明。目前我国房地产市场中所用的统

计指标一般都是房地产开发类的指标,是过去计划经济时代社会对固定资产投资状况统计量作为基础慢慢演变过来的。但随着社会经济的发展,房地产业已经有了翻天覆地的变化,除了一些基础的指标外很多指标的现实意义都不大。现行的房地产市场指标体系的建设已经远远落后与房地产本身的发展状况。这直接影响着普通市民对与我国房地产市场的正确认识。为此业内也在这一方面进行了积极的创造性研究。所以说在新的权威的统计指标得以规范之前,最好使用现行的最为基础的系统统计数据来进行预测。

其次目前现行的房地产市场统计指标不但与国际通用指标差异很大,而且国内各个地区所用的统计指标可能也不是完全一致的。这都会使得业内的专家学者对国内房地产状况研究和国际交流造成误解和不便。例如“人均居住面积”这个指标在国外是通过该地区的总面积与该地区的总人口作商就可以得出,但是在国内因为流动人口的原因就会有同一个指标不同部门统计出来的数据完全不同的情况出现;再说“空置面积”这个指标在国外是指所有空置房屋的面积(包括二手房和新房),而国内仅限于开发商新建住宅中未售出的商品房的面积。

最后政府机关以及行业部门等不同的单位可能对指标理解也有偏差,使得统计相同的指标可能有不同的数值。现在政府机关和行业部门之间统计范围和统计指标口径不一致的现象经常存在,这种现象在房地产统计系统中也很常见。例如:各级政府机关对房地产市场统计指标是按照国家统计局所制定的《房地产开发统计报表制度》,而建设部门系统统计的标准为建设部制定的《房地产统计报表制度》。这两种报表制度经常在不同的地方出现,这不单单使得相关统计数据意义不一样,也可能造成最终的数值有很大的差异。国家统计局 2006 年底天津市人均住房面积为 23.57 平方米,而天津市建设部 2006 年公布的天津市人均房地产建筑面积为 34.58 平方米。这种看似很诧异的现象,在房地产统计报表中却随处可见。这给读者和相关研究人员带来了很大的困惑。

综上所述,我国现行的房地产统计指标体系存在种种问题,而就是这些问题很大程度上导致相关责任部门不能利用正确的统计分析当前的房地产市场发展状况以及制定一系列切实可行的楼市宏观调控政策,严重影响着房地产市场健康稳定的发展。要建立切实可行的房地产市场预警体系就必须进一步完善房地产市场统计制度,严格执行房地产市场统计条例。所以说在没有解决房地产市场

所存在的各种问题的情况下,先解决对相关房地产市场基础统计数据得到的房地产市场预测模型较为合适。结合现行的房地产市场统计指标以及房地产市场本身固有特性,我们必须造出明确的、合适的、没有争议的指标来建立预测模型。

当然,考虑到市场因素与数据的准确性和易得性,本文研究的主要指标包括房地产开发投资额、房屋施工面积、房屋实际销售面积等,这些指标都是房地产市场发展水平的核心指标,研究的目的也就是对房地产市场中的主要指标参数进行模拟、分析与判断,从而得出模型,并给出最后的结论。

## **2.4 本章小结**

本章详细的阐述了房地产市场的基本概念和特征,解释了房地产市场的分类依据,与此同时总结了我国房地产业与我国国民经济之间的关系。在介绍完我国房地产市场基本情况后,详细的介绍了近些年青岛房地产市场发展水平,最后选取了本文的研究预测的对象为房地产开发投资额、房屋施工面积、房屋实际销售面积三个指标。

### 3 灰色预测模型

#### 3.1 灰色系统与灰色预测概述

科学界很多系统名称是由所研究的领域或者按创始人的姓名来命名的,但是有一种很常见的系统却是以颜色命名的,这就是——灰色系统。灰色系统最早是在控制理论领域中所提出来的。在控制领域中,系统的明确程度一般用颜色的深浅来表示:白色系统表示系统信息完全透明;黑色系统表示系统信息完全未知;而灰色系统是介于白色系统和黑色系统之间的一种状态,灰色系统是指系统信息有部分已经明确,部分未知的一种系统。灰色系统所研究的就是那些“不确定”、“贫信息”的系统,这些系统在日常生活和科学领域中十分常见,他们往往物理模型不确定、结构关系不明确、运行原理不清楚。而灰色系统的理论认为在这些物理模型不确定、结构关系不明确、运行原理不清楚的系统背后不然存在某种联系和规律,表象模糊离乱只是因为多方面因素干扰所造成的,这些系统本质上还是完整的、有序的。由此看来对灰色理论的研究具有很切实的现实意义。

灰色系统理论是有邓聚龙教授<sup>[37]</sup>在 1982 年首次提出后引起了国内外研究人员的持续广泛的关注。实践证明灰色系统理论在样本数据比较少、模型机制不明确的系统中的应用取得了良好的效果。灰色理论发展至今已经广泛应用与各个领域,从而被大家所熟知。

灰色系统理论的研究对象是那些“部分信息已知,部分信息未知”的“小样本”、“贫信息”不确定性系统,它主要通过从系统已经明确的信息中提取转换成能有价值能被利用的元素,进而对系统的运行、演化规律进行较为准确的描述和控制。一般来说灰色系统具有以下几个特点<sup>[38][39]</sup>:

- 1.灰色理论是常见的不确定系统方法。灰色系统主要研究对象就是那些不确定的系统,它把那些系统中不确定量看作灰度,通过一些数学方法处理不确定量使之得以量化。

- 2.灰色系统理论的核心就是通过已知信息处理探寻系统的发展规律,使之演化结果可模型化。灰色系统把灰色量处理得到时间序列,然后通过数学方法确定相关参数,最终得到数学模型。其中灰色预测并不是仅仅把相关数据看作一个随

机序列，而是把它看作是随着时间变化的灰色过程，它们之间是有某种内在的联系。

3.灰色系统理论能够处理那些“贫信息”不确定性系统。它与常规的统计模型不一样，它所需数据为短期的即可（一般要求超过 6 组）。因此对于那些只有少量样本的实验或者项目来说，灰色预测是一种十分简单有效方法

灰色模型 (Grey Model)简称  $GM$  模型，是灰色系统理论的基本模型，也是灰色控制理论的基础。灰色模型是从灰色系统理论中抽象提取出来的基本模型。它的研究内容是如何从一个信息量不是很充分的系统中提取相关有用信息并建立合理的数学模型，使得该模型能够为灰色系统的发展和演化提供合理准确的依据。

$GM(n,h)$  模型表示对  $h$  个变量用  $n$  阶微分方程建立的灰色模型。在  $GM(n,h)$  模型中，当  $h>2$  时， $GM(n,h)$  模型不能作为预测使用，一般只是用在系统因子之间的相互关系的分析中。作预测用的  $GM$  模型一般为  $GM(n,1)$  模型，其中最重要同时也是在日常科学研究中应用最为广泛的是运用灰色微分拟合法建立的  $GM(1,1)$  模型。

通常灰色预测是指以  $GM(1,1)$  模型为理论原型所进行的预测方法。 $GM(1,1)$  模型原理是通过对研究对象的原始时间序列进行再生成处理，弱化原始数据的随机误差，增强其数据的规律性，再通过建立一元微分形式建立模型，最终在通过还原方法处理得到预测值。从本质上说通过分析原始数据的趋势差异程度，并对其进行相关数学上的处理来探寻整个灰色系统变化规律，从而预测该系统对于未来变化的趋势。其中因为灰色模型具有所需的原始数据样本比较少，处理方法简单，使用范围广泛等优点，现在以及成为主流的预测方法之一。就  $GM(1,1)$  的运用范围来说的话，有自然灾害预测、国民经济预测、人口变化预测等等。特别是在缺乏统计数据资料的不确定性的系统中运用尤为广泛。但  $GM(1,1)$  预测并不是对于所有的不确定性系统都能适用，它在预测波动不是很大而且数据序列呈现指数变化的情况下效果较好，对于数据波动很大的数据序列可能得到的预测精度相对较低，因此在模型的应用过程中要根据实际情况对  $GM(1,1)$  进行改进，这样才

能达到预期的效果。

## 3.2 GM(1,1) 的建模过程

### 3.2.1 原始数据预处理

原始数据序列是系统特征的外在表现，也是预测的基础，预测归根究底就是通过对原始数据序列的分析，抽象归纳系统内在的规律，结合系统在一定条件下可能出现的变化情况，最终对系统的发展趋势进行判断预测。但是几乎所有现实生活中的系统都会被随机因素干扰，使其表现出来的特征并不准确，以至于一般来说数据序列并不是按照某种特定规律呈现的。如果仅仅使用未经过处理的原始数据进行预测往往会得到较差的精度。解决这一问题的常用办法就是建立模型之前要对相关原始数据进行预处理。运用灰色预测进行数据建模时，数据的预处理也是必不可少的一个过程。

数据的预处理是要遵循一定的规则，并且处理完以后要求得到的数据序列满足特定的条件。灰色预测要求经过处理的数据符合光滑或者准光滑条件，一般来说数据序列越是光滑，得到的预测精度就越高。灰色预测在实际运用过程中，如果原始数据序列已经符合光滑性标准了就可以直接使用模型了，否则就与要经过预测预处理使之达到使用条件，才能使用模型。

一般灰色预测要求模型有 90% 以上的精度，所以在建模之前要进行建模的可行性分析，也就是事前检验。事前检验要检验原始序列  $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$  中所有样本数据的级比  $\sigma(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}$  是否都落在区间  $(e^{-\frac{2}{n-1}}, e^{\frac{2}{n-1}})$  中。只有级比  $\sigma(k) \in (e^{-\frac{2}{n-1}}, e^{\frac{2}{n-1}})$  时，GM(1,1) 的精度才能达到要求，预测出来的数据才有意义。若  $\sigma(k) \notin (e^{-\frac{2}{n-1}}, e^{\frac{2}{n-1}})$ ，则该数据不能运用 GM(1,1) 模型，应该对原始数据进行数据预处理。数据预处理的原则是处理后的数据序列级比越接近 1 越好。

### 3.2.2 GM(1,1) 算法构造流程

目前使用最为广泛的灰色预测模型是一个变量、一阶微分的 GM(1,1) 模型。

下面简要的介绍一下模型 GM(1,1) 的含义：G ——表示灰色(Grey)；M ——表示模型(Model)；第一个“1” ——表示模型使用的是一阶方程；第二个“1” ——表示模型方程只含有一个变量。

GM(1,1) 模型的构造如下所述：

#### ①原始数据序列累加

原始数列  $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ ，其中， $x^{(0)}(n) \geq 0$ ， $k = 1, 2, \dots, n$ ，

$X^{(1)}$  为原始序列  $X^{(0)}$  经过一次累加生成序列： $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$ 。

其中  $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$ ， $k = 1, 2, \dots, n$ 。

#### ②建立灰色模型

构造背景值  $Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n))$ ，

其中  $z^{(1)}(k) = \alpha x^{(1)}(k-1) + (1-\alpha)x^{(1)}(k)$ ， $k = 2, 3, \dots, n$ ，一般取  $\alpha = 0.5$ 。

假定一次累加序列  $x^{(1)}$  具有近似指数规律。

对其建立一阶线性微分方程  $\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b$  3-1

称上式为 GM(1,1) 模型的白化方程，

#### ③模型参数求解

设方程的待估参数列为  $\hat{a} = [a, b]^T$ ，其中  $a$  称为发展系数，反映原始数据序列  $\hat{x}^{(0)}$  和  $\hat{x}^{(1)}$  的发展态势； $b$  为内生控制灰数，它是从背景值中挖掘出来的变量，反映数据之间变化的联系， $b$  是区别灰色建模和黑箱建模的标志。为了求解参数  $a$  和  $b$ ，令  $\hat{a} = [a, b]^T$  为待估向量，由于  $x^{(1)}(k)$  是离散的，将  $\frac{dx^{(1)}}{dt}$  离散化，则有



$$\frac{dx^{(1)}(k)}{dt} = x^{(1)}(k) - x^{(1)}(k-1), k = 2, 3, \dots, n$$

代入 3-1 式有

$$x^{(1)}(k) - x^{(1)}(k-1) + az^{(1)}(k) = b, k = 2, 3, \dots, n$$

$\hat{a}$  一般按照最小二乘法来求解, 易得  $\hat{a} = (B^T B)^{-1} B^T Y$

其中:

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix} \quad B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

#### ④构建预测公式

将求得的  $a$  和  $b$  代入 3-1 式, 求解该微分方程, 可得白化方程的解为

$$\hat{x}^{(1)}(k+1) = [x^{(1)}(0) - \frac{b}{a}]e^{-ak} + \frac{b}{a}, k = 1, 2, \dots, n. \quad 3-2$$

作一次累减, 可得模型预测值如下:

$$\hat{x}^{(0)}(1) = x^{(0)}(1), \quad \hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1), \quad k = 2, 3, \dots, n,$$

### 3.2.3 GM(1,1)模型的特征

灰色预测虽然具有模型机理简单、所需实验样本少、计算简便等一系列优点, 但是没有任何一种模型是万能的, 每种模型都有其适用的条件和范围,  $GM(1,1)$  模型也一样。灰色预测的适用范围与  $-a$  有很大关系, 一般来说只有  $|-a| < 2$  时  $GM(1,1)$  模型才有现实意义。通常情况下模型误差与  $a$  的关系很大, 相关文献已经说明了灰色预测模型参数与模型实用性具有下述关系<sup>[40]</sup>:

- ①  $-a \leq 0.3$  时,  $GM(1,1)$  用于中长期预测;
- ②  $0.3 < -a \leq 0.5$  时,  $GM(1,1)$  用于短期预测, 中长期预测谨慎使用;
- ③  $0.5 < -a \leq 0.8$  时, 用  $GM(1,1)$  用于短期预测也应谨慎;
- ④  $0.8 < -a \leq 1$  时, 用残差修正  $GM(1,1)$  模型;

⑤  $1 < -a$  时, 不宜采用  $GM(1,1)$  模型。

因此可以看出  $GM(1,1)$  是一种受到发展系数限制的长期预测模型, 而且一般来说最近的一两个预测数据可能具有良好的精度, 较远的数据更多只是反映一种趋势, 参考意义不是很大, 所以说对灰色模型作出改进在科学研究中的具有很现实的意义。

### 3.2.4 模型检验分析

如何才能得知所得到的预测值是精确的, 具有实际效果的呢? 这就需要对所得到的预测结果进行检验。

常用的检验方法包括:

①残差检验: 残差检验就是考察各个阶段的绝对误差和相对误差的大小, 得知模型精度的一种简单有效的检验方法。

残差(绝对误差)序列:  $\varepsilon^{(0)}(t) = x^{(0)}(t) - \hat{x}^{(0)}(t)$

相对误差序列:  $q^{(0)}(t) = (\varepsilon^{(0)}(t) / x^{(0)}(t)) \times 100\%$

平均残差:  $\varsigma = \frac{1}{n-1} \sum_{t=2}^n |q^{(0)}(t)|$

模型精度:  $\sigma = (1 - \varsigma) \times 100\%$

注:  $\sigma$  越大, 则说明模型的精度就越高, 模型的预测效果就越好。

②后验差检验

后验差检验是对于实际值与模型预测值之间的统计情况进行的模型精度检验方法。它是以残差(绝对误差)为基础, 由各阶段绝对误差绝对值的大小, 考察残差较小值点出现的概率情况, 与预测误差方差相关指标的大小。分别计算原始数据序列  $X^{(0)}$  与残差值序列  $\varepsilon^{(0)}$  的以下统计量:

实际值序列均值:  $\bar{X}^{(0)} = \frac{1}{n} \sum_{t=1}^n x^{(0)}(t)$

残差(绝对误差)序列均值:  $\bar{\varepsilon}^{(0)} = \frac{1}{n} \sum_{t=1}^n \varepsilon^{(0)}(t)$

$$\text{实际值序列均方差: } S_0^2 = \sum_{t=1}^n (x^{(0)}(t) - \bar{x}^{(0)}(t))^2, \quad S_0 = \sqrt{\frac{S_0^2}{n-1}}$$

$$\text{预测值序列均值: } \bar{\hat{X}}^{(0)} = \frac{1}{n} \sum_{t=1}^n \hat{x}^{(0)}(t)$$

$$\text{预测值序列均方差: } S_1^2 = \sum_{t=1}^n (\hat{x}^{(0)}(t) - \bar{\hat{x}}^{(0)}(t))^2, \quad S_1 = \sqrt{\frac{S_1^2}{n-1}}$$

$$\text{残差(绝对误差)序列均方差: } S_2^2 = \sum_{t=1}^n (\varepsilon^{(0)}(t) - \bar{\varepsilon}^{(0)}(t))^2, \quad S_2 = \sqrt{\frac{S_2^2}{n-1}}$$

$$\text{均方差比(后验差系数)值: } C = \frac{S_2}{S_0}$$

$$\text{小误差概率: } p = P\{|\varepsilon^{(0)}(t) - \bar{\varepsilon}^{(0)}| < 0.6745S_0\}$$

表 3-1  $GM(1,1)$  模型预测精度等级划分标准<sup>[41]</sup>

相对误差 $q^{(0)}$	均方差比值 $C$	小误差概率 $p$	精度等级
0.01	0.35	0.95	一级
0.05	0.5	0.8	二级
0.1	0.65	0.7	三级
0.2	0.8	0.6	四级

由表 3-1 可以可知,  $q^{(0)}$  和  $\varepsilon^{(0)}$  都要求越小越好, 两者相差越小表示预测值和实际值越接近;  $C$  越小越好,  $C$  小表示  $S_0$  与  $S_2$  相比较小, 即说明  $X^{(0)}$  离散程度大,  $\varepsilon^{(0)}(t)$  离散性小。  $p$  值越大越好,  $p$  大表明误差较大的概率小, 表示模型精度较高。若  $q^{(0)}$ 、 $C$  和  $p$  都在允许条件范围之内时, 可以用该灰色模型来进行预测。否则应对该模型进行适当的改进, 直到通过上述检验为止。

### 3.3 $GM(1,1)$ 模型的局限性及改进

#### 3.3.1 $GM(1,1)$ 模型的局限性

$GM(1,1)$  模型对于波动性很强的时间序列预测效果一般较差, 因为  $GM(1,1)$

模型仅仅是根据历史数据预测的而没有考虑各因素之间的关系,所以一般来说灰色预测对于中长期预测往往模型精度比较差。另外我们可以分析  $GM(1,1)$  模型建模过程,发现存在一些不合理的地方<sup>[42]</sup>。

第一,在  $GM(1,1)$  模型求解过程中,背景值是有原始数列经过一次累加生产的,一般我们取权值  $\alpha$  为 0.5。在理论上没有任何依据可以说明取  $\alpha = 0.5$ , 这时的模型精度是最高的。业内很多学者通过对  $\alpha$  的设定进行改进来提高模型的精度。有很大相关实例验证了通过这种方式提高模型精度是有效的。如果仅仅是对  $\alpha$  的取值进行改进而不进行其他的优化,往往并不能达到预期的效果。本文除了对背景值进行了优化,而且对模型的初始值进行了改进,实验证明取得了良好的效果。

第二,模型将  $X^{(0)}(1)$  作为求解式子 3-1 的初始值,即假定使得  $\hat{X}^{(1)}(1) = X^{(0)}(1)$ , 得到的拟合曲线  $\hat{X}^{(1)}$  在坐标平面  $(k, \hat{X}^{(1)}(k))$  上也必然通过点  $(1, X^{(0)}(1))$ 。然而,由于灰色预测  $GM(1,1)$  模型从本质来说的化应该使得结果应满足  $\sum_{k=1}^n (X^{(0)}(k) - \hat{X}^{(0)}(k))^2$  最小。根据最小二乘法原理,拟合曲线并不一定通过第一个数据点。所以说把  $\hat{X}^{(1)}(1) = X^{(0)}(1)$  并没有理论上的依据。近些年很多学者对灰色预测初始值的选取进行了研究。罗荣桂,陈炜认为应该选择最后一个数据  $X^{(0)}(n)$  作为初始值,因为  $X^{(0)}(n)$  是最近更新的,所包含的信息往往是更为准确的。张大海等认为,应将原始数据序列中的所有数据分别作为模型的初始值进行计算模型的精度,选择精度最高的  $X^{(0)}(k)$  作为模型的初始值进行预测。但是由于各个方面的因素的干扰导致于原始数据序列必然存在随机误差,理论上,使预测模型精度最高的初始值可能并不存在  $\{X^{(0)}(k)\}$  中,所以把原始数据序列中的一个值作为初始值不能从理论上保证模型的精度是最高。

### 3.3.2 $GM(1,1)$ 模型的改进

目前  $GM(1,1)$  模型的改进途径最主要在以下几个方面:

第一、对初始值进行修正。在常见的  $GM(1,1)$  模型中往往是通过采取原始数据序列中第一个数据作为初始值进行建模的,这样的做法并不是最为合理的。直接选

取第一个数据作为初始值肯定会造成一定的系统误差。所以说对初始值进行修正在一定程度上可以提高模型精度。

第二、改进模型对时间响应函数的选取。在常见的  $GM(1,1)$  模型中初始值对时间响应函数的选定影响很大，如果能找到与初始值没有关系的时间响应函数，必然可以减少系统误差，提高模型精度。

第三、通过  $GM(1,1)$  模型的残差分析建立新型的  $GM(1,1)$  模型。我们可以对  $GM(1,1)$  模型的残差建立模型，并通过此模型与  $GM(1,1)$  模型相结合建立新的模型，这是关于  $GM(1,1)$  模型研究的重要方面。

第四、运用处理好的原始数据建立预测模型。对于那些波动性很大，不能呈现相关指数性质的原始数据序列，如果直接运用  $GM(1,1)$  模型进行预测，一般得到的预测效果较差。所以通过相关数据处理，增强原始数据序列的光滑度和规律性，这也是提高灰色预测的一条很重要的途径。

第五、寻找最优的背景值参数或者改进传统  $GM(1,1)$  模型背景值构造的方法。一般来说我们常取  $\alpha x^{(1)}(k-1) + (1-\alpha)x^{(1)}(k)$  作为  $GM(1,1)$  模型的背景值，它通过梯形公式来构造的，我们可以通过参数寻优的方式找到最佳的  $\alpha$ ，另外我们也可以使用其他更为精确的背景值构造方式，也能很有效的降低误差，提高模型精度。

第六、直接寻求最优模型参数  $a$  和  $b$ 。平时我们都是通过最小二乘法得到模型参数  $a$  和  $b$ ，但就必然涉及到背景值的构造问题。如果能够避开背景值的选取，使用其他方法直接寻求最优模型参数  $a$  和  $b$ ，也不为是提高模型精度的一种好方法。

### 3.4 改进 $GM(1,1)$ ——马尔柯夫模型

#### 3.4.1 基于背景值和灰色作用量的 $GM(1,1)$ 模型改进

经过上述分析研究可知  $GM(1,1)$  模型产生系统误差的主要原因是：当发展系数  $\alpha$  的绝对值 ( $|\alpha| > 0.5$ ) 较大时，使得模型精度较差，得到的中长期预测的结果往

往没有实际意义, 基于模型的初值和背景值的改进在很大程度上会使得模型的发展系数得以优化, 通过这种方式改进的灰色预测模型不但保留这传统  $GM(1,1)$  模型的优点, 而且一般对于中长期预测精度也有很大的改善。

下面将结合优化初值和优化背景值对  $GM(1,1)$  模型进行改进。其建模过程如下:

#### ①原始数据序列累加

设  $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ , 其中,  $x^{(0)}(n) \geq 0$ ,  $k = 1, 2, \dots, n$ ,  $X^{(1)}$  为  $X^{(0)}$

的一次累加生成序列:  $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$ ,

其中  $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$ ,  $k = 1, 2, \dots, n$ 。

#### ②建立灰色模型

构造背景值  $Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n))$ , 其中,

$$z^{(1)}(k) = \frac{x^{(1)}(k) - x^{(1)}(k-1)}{\ln x^{(1)}(k) - \ln x^{(1)}(k-1)}, \quad k = 2, 3, \dots, n,$$

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b_1 + b_2 t \quad 3-3$$

称 3-3 式为改进型  $GM(1,1)$  模型的白化方程

#### ③模型参数求解

设待估参数列  $\hat{a} = [a, b_1, b_2]^T$ ,

还是按照最小二乘法来求解模型参数, 易知  $\hat{a} = (B^T B)^{-1} B^T Y$

$$\text{其中: } Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix} \quad B = \begin{bmatrix} -z^{(1)}(2) & 1 & 2 \\ -z^{(1)}(3) & 1 & 3 \\ \dots & \dots & \dots \\ -z^{(1)}(n) & 1 & n \end{bmatrix}$$

#### ④构建预测公式

因为一阶非齐次线性方程  $\frac{dy}{dx} + p(x)y = f(x)$  的通解为

$$y = e^{-\int p(x)dx} (c + \int f(x)e^{\int p(x)dx} dx); \quad 3-4$$

其中  $c$  为任意常数

所以 3-3 的通解为:

$$x^{(1)}(t) = e^{-\int a dt} \left( \int (b_1 + b_2 t) e^{\int a dt} dt + c \right) = \frac{b_1}{a} + \frac{b_2}{a} t - \frac{b_2}{a^2} + c e^{-at}; \quad 3-5$$

$c$  为任意常数

由  $x^{(1)}(t) \Big|_{t=1} = x^{(1)}(1)$  可以得到

$$x^{(1)}(1) = \frac{b_1}{a} + \frac{b_2}{a} - \frac{b_2}{a^2} + c e^{-a}$$

$$c = (x^{(1)}(1) - \frac{b_1}{a} - \frac{b_2}{a} + \frac{b_2}{a^2}) e^a$$

$$\text{可得 3-3 的解为 } x^{(1)}(k+1) = (x^{(0)}(1) - \frac{b_1}{a} - \frac{b_2}{a} + \frac{b_2}{a^2}) e^{-ak} + \frac{b_1}{a} + \frac{b_2}{a} (k+1) - \frac{b_2}{a^2}$$

作一次累减, 可得模型预测值如下:

$$\hat{x}^{(0)}(1) = x^{(0)}(1)$$

$$x^{(0)}(k+1) = x^{(1)}(k+1) - x^{(1)}(k) = (1 - e^{-a})(x^{(0)}(1) - \frac{b_1}{a} - \frac{b_2}{a} + \frac{b_2}{a^2}) e^{-ak} + \frac{b_2}{a}, k = 1, 2, \dots, n-1 \quad 3-6$$

我们很容易可以从 3-6 中看出模型的精度除了取决于常数  $a, b_1$  和  $b_2$  以外, 而且

$x^{(0)}(1)$  的选取同样对该模型的预测精度有很大的影响。 $GM(1,1)$  模型的初值一般选取原始数据序列的第一个数据值是没有理论依据的, 在相关文献实验中也证明这么做并不合理。因此如何选择初值提高模型的预测精度也是有待我们解决的问题。本文在寻取最小化指标函数的情况下, 采用如下方法对  $x^{(0)}(1)$  进行了修正。

设模型的初值为  $\lambda x^{(0)}(1)$ , 可以得到

$$x^{(1)}(k+1) = (\lambda x^{(0)}(1) - \frac{b_1}{a} - \frac{b_2}{a} + \frac{b_2}{a^2}) e^{-ak} + \frac{b_1}{a} + \frac{b_2}{a} (k+1) - \frac{b_2}{a^2}, \quad k = 1, 2, \dots, n-1$$

$$\text{建立目标函数 } J = \min \sum_{k=2}^n [\hat{x}^{(1)}(k) - x^{(1)}(k)]^2$$

采用如下梯度法求取指标函数最小值时的待定参数,

$$\begin{aligned} \frac{dJ}{d\lambda} &= \sum_{k=2}^n [(\lambda x^{(0)}(1) - \frac{b_1}{a} - \frac{b_2}{a} + \frac{b_2}{a^2}) e^{-2a(k-1)} + \frac{b_1}{a} + \frac{b_2}{a} k - \frac{b_2}{a^2} - x^{(1)}(k)] x^{(0)}(1) e^{-a(k-1)} \\ &= \sum_{k=2}^n \lambda x^{(0)}(1) e^{-2a(k-1)} - \frac{b_1 + b_2}{a} \sum_{k=2}^n e^{-2a(k-1)} + \frac{b_2}{a^2} \sum_{k=2}^n e^{-2a(k-1)} + (\frac{b_1}{a} - \frac{b_2}{a^2}) \sum_{k=2}^n e^{-a(k-1)} + \frac{b_2}{a} \sum_{k=2}^n e^{-a(k-1)} - \sum_{k=2}^n x^{(1)}(k) e^{-a(k-1)} = 0 \end{aligned}$$

$$\text{令 } d = \sum_{k=2}^n e^{-2a(k-1)}, f = \sum_{k=2}^n e^{-a(k-1)}, g = \sum_{k=2}^n k e^{-a(k-1)}, h = \sum_{k=2}^n x^{(1)}(k) e^{-a(k-1)}$$

$$\text{则 } \lambda x^{(0)}(1) d - (\frac{b_1 + b_2}{a} - \frac{b_2}{a^2}) d + (\frac{b_1}{a} - \frac{b_2}{a^2}) f + \frac{b_2}{a} g - h = 0$$

$$\text{得 } \lambda = [(\frac{b_1+b_2}{a} - \frac{b_2}{a^2})d + (\frac{b_1}{a} - \frac{b_2}{a^2})f - \frac{b_2}{a}g + h] / x^{(0)}(1)d \quad 3-7$$

### 3.4.2 马尔柯夫模型

#### (1) 马尔柯夫状态划分

利用马尔柯夫链对系统作出预测的关键在于了解系统状态的转移状况。一般来说首先对所有的数据序列进行状态的划分(把已有的数据分别归类为若干个状态),  $Y(t)$  作为马尔科夫链状态划分的基准, 把  $Y(t)$  划分成若干个条形区域, 每个区域就是一种状态。根据已知数据落入状态区域的情况就会得到状态概率转移矩阵<sup>[43]</sup>。下面我就简单的介绍以下模型建立的流程。

对于符合马尔科夫链特点的随机数列可根据具体的情况划分为  $n$  个状态, 其中状态  $\theta_i$  可表示为:

$$\theta_i = [\theta_{1i}, \theta_{2i}] \quad \theta_{2i} \leftarrow \theta_{1i}$$

$$\theta_{1i} = Y(t) + A_i \quad (i = 1, 2, \dots, n) \quad 3-8$$

$$\theta_{2i} = Y(t) + B_i \quad (i = 1, 2, \dots, n) \quad 3-9$$

由于  $Y(t)$  是关于时间  $t$  的函数, 因而状态边界  $\theta_{1i}$ ,  $\theta_{2i}$  也随着时间序列变化而变化的。即状态  $\theta_i$  具有动态性。关于  $\theta_i$  的含义、状态划分数目  $n$  和状态边界  $\theta_{1i}$ ,  $\theta_{2i}$  的确定, 可根据原始数据数目来确定。

#### (2) 计算状态转移概率矩阵

$$P_{ij} = M_{ij} / M_i \quad (i, j = 1, 2, \dots, n) \quad 3-10$$

式中  $M_{ij}$  为由状态  $\theta_i$  经过  $m$  步转移到状态  $\theta_j$  的原始数据样本个数;  $M_i$  为处于状态  $\theta_i$  的原始数据样本数。



$$\text{状态转移概率矩阵 } R(m) = \begin{pmatrix} P_{11}(m) & P_{21}(m) & \cdots & P_{n1}(m) \\ P_{21}(m) & P_{22}(m) & \cdots & P_{2n}(m) \\ \cdots & \cdots & \cdots & \cdots \\ P_{n1}(m) & P_{n2}(m) & \cdots & P_{nn}(m) \end{pmatrix}$$

$R(m)$  反映了系统中各状态之间转移的规律。其中  $P_{ij}(m)$  表示由状态  $\theta_i$  经过  $m$  步转移到状态  $\theta_j$  的概率。 $R(m)$  是马尔科夫模型预测的基础。通过观察  $R(m)$  的则可预测系统的变化情况。一般情况下我们只考察一步转移概率矩阵  $R(1)$  中的第  $k$  行, 如果有  $\max_j P_{kj}(1) = P_{kl}(1)$ , 则表明下一时刻系统最可能由  $\theta_k$  状态转向  $\theta_l$  状态。

### (3) 确定预测值的变动区间和预测值

通过考察  $R(1)$ , 确定了系统下一状态后, 也就确定了预测值的变动区间为  $[\theta_{li}, \theta_{2i}]$ 。

最有可能的预测值  $Y(t)$  预测值可由下式计算:  $\hat{Y}(t) = (\theta_{li} + \theta_{2i}) / 2$  **3-11**

由 3-8、3-9、3-10 得:  $\hat{Y}(t) = Y(t) + (A_i + B_i) / 2$  **3-12**

## 3.5 改进 $GM(1,1)$ ——马尔柯夫模型的应用

要使用灰色预测对未来青岛房地产市场相关指标进行预测研究的前提是, 它是一个灰色系统, 符合使用灰色预测的所有条件。灰色理论把系统的随机过程看成一个灰色过程, 透过现象看本质, 它将随机性比较强的序列通过转换生成规律性比较强的数据序列, 从而可以建立微分方程进行高精度模拟。房地产市场可以看成一个灰色系统, 与无数的可预知和不可预知的外部因素影响它, 使其显示出来的指标特征有很大的随机性和波动性, 由于它的复杂性和不规律性, 需要找到一种预测模型对其进行准确的预测就显得不是很容易。灰色预测不要太多的样本数据就可以建立高精度预测模型, 而且模型使用条件较为宽泛易于达到, 所以应用灰色预测对青岛市房地产市场预测就比较合适。

根据相关的灰色建模理论, 青岛市房地产市场系统符合  $GM(1,1)$  预测模型的条件如下:

- ①房地产市场可以看成是一个灰色系统，其相关指标的变动可以看作是灰色量；
- ②本文分析所采用的房地产相关指标的原始数据大于零，具有现实的含义；
- ③本文分析所采用的房地产相关指标的原始数据无强烈波动，是等间距的，时间间隔为 1 年。
- ④本文分析所采用的房地产市场数据样本满足灰色  $GM(1,1)$  预测模型对于系统的最低数据要求，即  $n > 4$ 。

### 3.5.1 改进 $GM(1,1)$ ——马尔柯夫模型建立

#### 1) 数据

本文中原始数据的获取是以历年《青岛统计年鉴》为基础，结合青岛市统计局、青岛市土地与房地产管理局所公开发布(含网站上公布)的数据。在搜寻整理这些数据的过程中，发现有有些数据不完全、不系统并与市场实际数据存在差异等一些问题，所以本文在这里选取了从 1997 年到 2010 年的数据作为研究依据，2011-2012 年的数据作为检验样本数据序列，预测 2013-2016 年的房地产市场相关参数指标。本节以房屋施工面积为预测对象进行说明，房地产开发投资额、房屋实际销售面积这两个指标的计算过程类似。

**表 3-2 1997-2012 年房屋施工面积、房地产开发投资额、房屋实际销售面积**

年 份	1997 年	1998 年	1999 年	2000 年	2001 年	2002 年	2003 年	2004 年
房屋施工面积(万平方米)	675	742	900	1047	1374	1417	1765	2101
房地产开发投资额(亿元)	42.66	43.81	62.97	67.51	92.52	103.65	127.80	162.70
房屋实际销售面积(万平方米)	113	178	240	301	406	427	469	516
年 份	2005 年	2006 年	2007 年	2008 年	2009 年	2010 年	2011 年	2012 年
房屋施工面积	2363	2751	3224	3773	4310	5058	5693	6473
房地产开发投资额(亿元)	223.84	268.36	322.35	380.56	459.48	602.43	782.71	930.10
房屋实际销售面积(万平方米)	740	719	833	770	1262	1360	1026	950.9

本文把改进  $GM(1,1)$  ——马尔柯夫模型通过编程输入 Matlab<sup>[44]</sup>程序中以便可以快捷、准确的得到相关预测结果：

#### 2) 改进型灰色 $GM(1,1)$ 建立模型及检验

①级比检验，进行建模可行性分析

房屋施工面积原始数据序列为  $X$ ，则有

$$X = (675, 742, 900, 1047, 1374, 1417, 1765, 2101, 2363, 2751, 3224, 3773, 4310, 5058)$$

原始数据的级比序列为

$$\sigma(k) = (0.91, 0.82, 0.86, 0.76, 0.97, 0.80, 0.84, 0.89, 0.86, 0.85, 0.85, 0.88, 0.85)$$

14 个原始数据所对应的级比覆盖区域为  $\sigma(k) \in (0.875, 1.143)$ ，而原数据中有 9 个样本的级比没有落入该区间，因此不能用没有处理过的原始数据进行  $GM(1,1)$  建模。

②原始数据处理

为改进原始数据序列  $X^{(0)}$  的级比，本文采取对  $X^{(0)}$  作方根变换的方法。令方根变换值为  $X^{(0)}(k) = \sqrt{X(k)}, k = 1, 2, \dots, 14$ ，得到方根序列  $X^{(0)}(k)$

$$X^{(0)} = (25.98, 27.24, 30.00, 32.36, 37.07, 37.64, 42.01, 45.84, 48.61, 52.45, 56.78, 61.42, 65.65, 71.12)$$

处理过的数据级比序列为

$$\sigma(k) = (0.95, 0.91, 0.93, 0.87, 0.98, 0.90, 0.92, 0.94, 0.93, 0.92, 0.92, 0.94, 0.92)$$

除  $\sigma(4)=0.87$  外， $\sigma(k)$  中其余各点级比均符合条件，因此  $X^{(0)}$  可使用  $GM(1,1)$  模型建模。

③改进型灰色预测建模过程

$$\text{一次累加生成序列: } X^{(1)} = (25.98, 53.22, 83.22, 115.58, 152.65, 190.29, 232.30, 278.14, 326.75, 379.20, 435.98, 497.40, 563.05, 634.17)$$

$$\text{背景值 } Z^{(1)} = (37.99, 67.11, 98.52, 133.26, 170.78, 210.60, 254.53, 301.79, 352.32, 406.93, 466.02, 529.55, 597.91)$$

$$\text{待估参数列 } \hat{a} = [a, b_1, b_2]^T = (B^T B)^{-1} B^T Y = (-0.0610, 23.6268, 0.7659)$$

修正后的数据初值系数为  $\lambda = 1.0853$

$$\begin{aligned} \text{预测序列 } x^{(0)}(k+1) &= x^{(1)}(k+1) - x^{(1)}(k) = (1 - e^a) \left( \lambda x^{(0)}(1) - \frac{b_1}{a} - \frac{b_2}{a} + \frac{b_1}{a^2} \right) e^{-ak} + \frac{b_2}{a} \\ &= x^{(0)}(k+1) = 37.5055 * e^{0.061k} - 12.5481 \end{aligned}$$

=(25.98 ,29.53 ,29.83 ,32.49 ,35.33 ,38.34 ,41.54 ,44.95 ,48.57 ,52.41 ,56.50 ,  
60.85 ,65.47 ,70.38 ,75.60 ,81.14 ,87.04 ,93.31 ,99.97 ,107.05)

还原预测值:  $\hat{X}^{(0)} = (674.96 ,872.27 ,889.64 ,1055.85 ,1248.11 ,1470.09 ,1725.94 ,$   
 $2020.40 ,2358.80 ,2747.21 ,3192.51 ,3702.48 ,4285.96 ,4952.96 ,5714.79 ,6584.30 ,$   
 $7576.01 ,8706.37 ,9994.02 ,11460.03 )$

#### ④模型预测检验

残差检验:

将  $x^{(0)}(k)$  及其预测值  $\hat{x}^{(0)}(k)$ 、残差  $\Delta x^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k)$ 、相对误差:

$q^{(0)} = \frac{\Delta x^{(0)}(k)}{x^{(0)}(k)}$  的计算结果列表, 如表 4-2 所示。

表 3-3 改进型灰色预测模型残差检验值表

年份	原始值	预测值	残差	相对误差
1998	742	872.27	-130.27	-17.56%
1999	900	889.64	10.36	1.15%
2000	1047	1055.85	-8.85	-0.84%
2001	1374	1248.11	125.89	9.16%
2002	1417	1470.09	-53.09	-3.75%
2003	1765	1725.94	39.06	2.21%
2004	2101	2020.40	80.60	3.84%
2005	2363	2358.80	4.20	0.18%
2006	2751	2747.21	3.79	0.14%
2007	3224	3192.51	31.49	0.98%
2008	3773	3702.48	70.52	1.87%
2009	4310	4285.96	24.04	0.56%
2010	5058	4952.96	105.04	2.08%

模型的平均残差为: 3.41%;模型的预测精度为 96.59%>90%, 所以该模型对房屋  
施工面积这个指标的预测效果很好。

后验差检验:

初始值序列  $x$  的均值和方差分别为  $\bar{x} = 2371.15$ ;  $S_0^2 = 1748622$ ;  $S_0 = 1322.3547$

残差序列  $\Delta$  的均值和方差分别为  $\bar{\Delta} = 52.86$ ;  $S_2^2 = 1932.89$ ;  $S_2 = 43.96$

则后验差比值为  $C = \frac{S_2}{S_0} = 0.0332 < 0.35$ ;

经过残差检验和后验差检验，小误差概率为  $p = 1 > 0.95$

$q^{(0)}$ 、 $C$  和  $p$  参照表 3-1 对照，上述所建立的改进  $GM(1,1)$  模型，预测精度等级为一级(好)。

### (3) 马尔柯夫模型建立模型及检验

根据灰色预测得到的预测值，综合考虑模型的易操作性和准确性，划分数据样本为如下四个状态：

$\theta_1$	$\theta_{11} = x^{(0)}(k) - 150$	$\theta_{21} = x^{(0)}(k) - 75$
$\theta_2$	$\theta_{12} = x^{(0)}(k) - 75$	$\theta_{22} = x^{(0)}(k)$
$\theta_3$	$\theta_{13} = x^{(0)}(k)$	$\theta_{23} = x^{(0)}(k) + 80$
$\theta_4$	$\theta_{14} = x^{(0)}(k) + 80$	$\theta_{24} = x^{(0)}(k) + 150$

式中  $x^{(0)}(k)$  为  $k$  时刻按照上述灰色模型得出的预测值，将原始数据、预测值曲线和四个划分状态绘制成图 3-1，

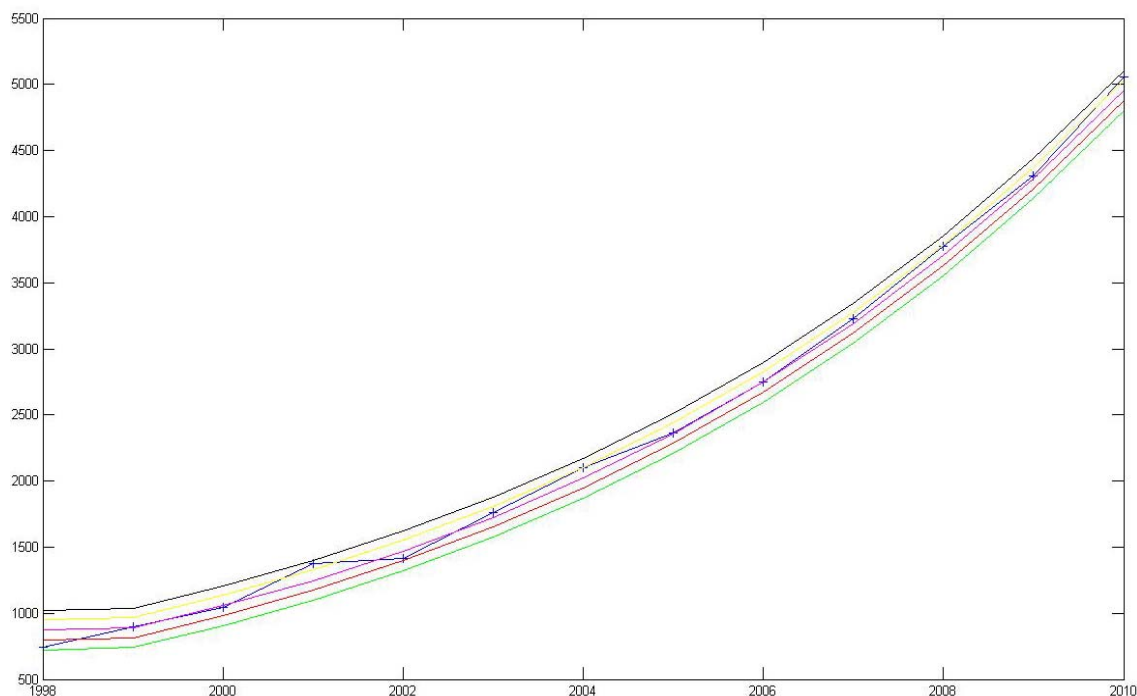


图 3-1 马尔科夫链状态划分

由图 3-1 可知，落入  $\theta_1, \theta_2, \theta_3, \theta_4$  状态的原始数据样本点数为  $L_1=1, L_2=2, L_3=7,$

$L_4=3$ , 由状态  $\theta_1$  一步转移到状态  $\theta_1, \theta_2, \theta_3, \theta_4$  的原始数据样本个数分别为  $L_{11}=0$ ,  $L_{12}=0$ ,  $L_{13}=1$ ,  $L_{14}=0$ 。依次类推, 可计算  $L_{ij}(i=1,2,3,4; j=1,2,3,4)$  的值。根据 3-10 可计算出系统各个状态的一步转移概率为  $P_{ij}(1)(i, j=1,2,3,4)$ , 所以最终构成的状态转移概率矩阵如下:

$$R(1) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{7} & 0 & \frac{4}{7} & \frac{2}{7} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 \end{pmatrix}$$

根据以上矩阵可预测 2011 房屋施工面积。由于图 1 可以看出 2010 年房屋施工面积落于  $\theta_4$  状态, 则考察矩阵第四行有  $\max P_{4j} = P_{41} = P_{43} = \frac{1}{3}$ 。因此, 预测 2011 年房屋施工面积最有可能处于  $\theta_3$  或者  $\theta_1$  状态 (这里取  $\theta_1$  状态), 则房屋施工面积最有可能在  $[\hat{x}(15)^{(0)}-150, \hat{x}(15)^{(0)}-75]$ , 即  $[5564.79, 5639.79]$ , 即 2011 年房屋施工面积比较有可能的预测值为:  $y(15) = (\theta_{11} + \theta_{21}) / 2 = 5602.29$

同理 2012 年房屋施工面积的预测值为  $y(16) = (\theta_{11} + \theta_{21}) / 2 = 5602.29$

**表 3-4 2011-2012 年房屋施工面积预测值参差检验值表**

年份	原始值	预测值	残差	相对误差
2011	5693	5602.29	90.71	1.59%
2012	6473	6471.79	1.21	0.02%

从表 3-4 中可以看出运用该模型得到 2011、2012 年预测值与真实值误差很小, 所以说该模型对房屋施工面积进行预测是非常合适和有效的。

### 3.5.2 实验结果分析

按上述步骤可以得到各指标的 2011-2016 年的预测值如下表所示

表 3-5 2011-2016 年青岛房地产市场发展状况预测表

年份	2011	2012	2013	2014	2015	2016
房屋施工面积	5602.29	6471.79	7616.01	8706.37	9994.02	11460.03
房地产开发投资额	718.94	881.61	1080.36	1323.11	1619.50	1981.30
房屋实际销售面积	<b>1461.21</b>	<b>1668.74</b>	<b>1902.56</b>	<b>2165.79</b>	<b>2461.89</b>	<b>2794.76</b>

注：因为本文用的是一步马尔科夫转移模型进行预测的，所以 2011-2013 年的数据可以运用改进  $GM(1,1)$ ——马尔柯夫模型进行预测的，而 2014-2016 年的预测数据仅为改进  $GM(1,1)$  模型进行预测得到的。

其中：房地产开发投资额的预测值通过了模型检验，证明能用上述方法进行计算和预测的。但是房屋实际销售面积指标由于在建模过程中所得预测精度为  $89.87\% < 90\%$ 。而且在对 2011-2012 年该数据进行比较发现误差超过了 25%，所以该指标不适用上述模型进行预测。

从上述实验结果可以看出并不是所有符合灰色理论的数据序列都能用灰色方法来预测，所以我们必须重新找到另一种合适有效的方法来进行模拟，接下来的一章就是用比较流行的预测方法——神经网络来对这三组指标进行预测。

### 3.6 本章小结

本章详细介绍了传统  $GM(1,1)$  建模思想和流程，指出其存在的不合理的地方，并给出相关的解决方案。给出了基于改进背景值和灰色作用量的新型  $GM(1,1)$  模型。通过此模型和马尔科夫模型的结合使用，预测了青岛房地产市场的相关指标，取得了良好的实际效果。

## 4 神经网络预测模型

### 4.1 神经网络简介

从 19 世纪初开始预测理论特别是线性预测理论方法得到了飞速的发展，然而随着人们对自然界不断深入的了解，越来越多的现象并不能通过线性预测得以解决，线性预测的方法已经远远不能够满足科学研究的需要了，近些年非线性预测已经称为复杂大系统预测分析的主要研究方向了。简单系统的非线性预测问题一般通过常见的非线性预测方法就能解决，然而对于大型的复杂的系统常规的非线性方法就往往不太适用。所以急需寻找到一种处理这些大型复杂问题的非线性方法，于是神经网络技术就在这种背景中诞生了。

神经网络也被称为人工神经网络（Artificial Neural Network），简称 ANN。它是模拟人类大脑神经网络结构和功能的从而建立的能实现信息处理的数学模型<sup>[45]</sup>。它有着很强的自适应非线性处理和容错能力，理论上神经网络可以以任意精度逼近所有的非线性函数，所以在优化控制、趋势预测以及诊断辨别方面它有着得天独厚的优势。简单的说，ANN 就是连接着输入和输出神经元的一种关系，关系的紧密程度在数学上就用权值的大小表示。权值的大小是通过学习样本训练网络得到的，在神经网络的训练学习过程就是不断调整网络中各个权值，使得输入值可以通过网络能够得到的符合精度的预测值。总的说来就是不断优化人工神经网络中各个神经元之间的关系，从而达到我们所需要的目的。

ANN 具有以下特征<sup>[46]</sup>：

①ANN 的信息特征分布在其网络结构中的神经元和神经元之间的关系上的。这也就是说 ANN 的信息存储模式是分布式的，它的信息不再某个位置或者节点上，而是存在与整个网络中。

②ANN 和人脑一样具有并行处理信息的能力。其表现在：各个神经元都是独立处理信息的小系统，并不接受上级或者下级的神经元的控制，它具有独立处理信息的能力，接受其他神经单元的输出经过处理后反馈到网络中去。

③ANN 各层之间的连接关系是全连接关系，这使得 ANN 具有超强的自组织性和容错能力，整个网络会自动舍弃无效的神经元重新组合网络（无效的神经元与外界的关系权值变为 0），从而使得局部的网络结构不会对整体的网络结构造



成很大的影响。

④ANN 处理信息的速度很快。这是因为它是有相当数量的神经元相互并行处理的造成的。每个神经元虽说功能简单但都有独立处理信息的能力，大量的神经元集体处理信息必然会有较快的速度。

⑤ANN 是适应能力非常强大。在复杂的 ANN 结构中，结构和权值都在不断的训练中不断优化，直至样本的输出经过网络得到的输出符合样本的输出，在这个过程中就形成了一个符合实际意义的 ANN 结构，这个过程类似与人类的学习过程，通过学习规律性的东西，输出符合规律的结果。

⑥ANN 可以逼近任意非线性函数，一般的网络结构中，有一个或者多个隐含层，在每一个隐含层中都建立的不同的关系函数，这使得网络本身就能自动的逼近数据样本所反映的规律，从而使得网络有着很强的非线性模拟能力。

⑦ANN 广泛的应用于各个领域，使得那些背景复杂、机理模糊、现象随机的问题都找到了一种得以处理的方法。

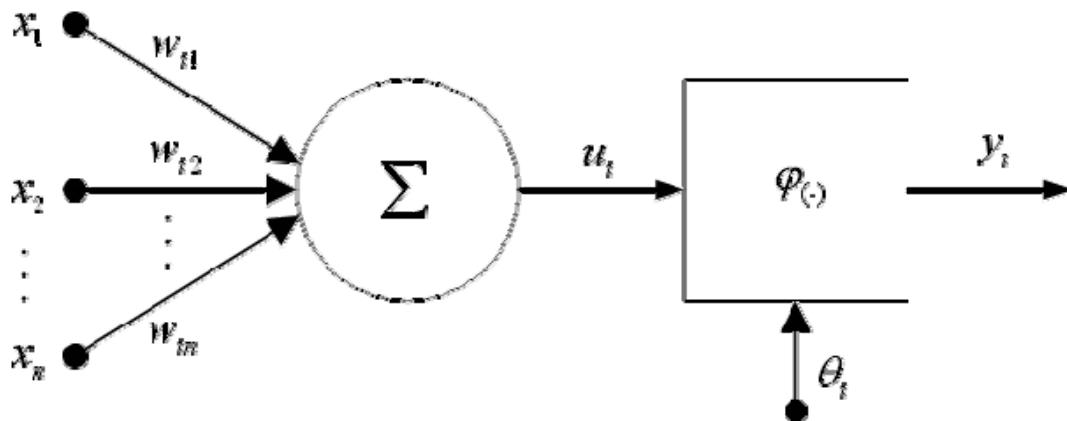
ANN 可以在不寻求具体的数学模型，就能够模拟逼近任意非线性函数的特性，这使得神经网络称为现代非线性预测方法研究的热点，自从 1980 年神经网络概念被提出后，吸引的各个领域的科学家对它进行研究和应用，也使得神经网络的理论在逐步的完善和发展，到 20 世纪以来，神经网络已成为经济预测、灾害预测、能源消耗预测等一系列社会经济问题预测的重要方法。越来越多的实验研究结论表明，在复杂的社会系统研究预测方法中，ANN 明显优于指数平滑、趋势外推、ARIMA 等方法，而且系统越是复杂、非线性形态越是显著，它的优势就越是明显。神经网络理论的发展不光是推动了预测学的发展，而且带动着一系列相关学科的发展，比如：智能控制、图像识别、故障诊断、系统辨识等等，并取得了卓越的成果。经过数十年的研究和发展，ANN 已经形成了数十种成熟的网络结构模型。常见的有：前向反馈神经网络（Feedforward NNs）、向后反馈神经网络（Feedback NNs）和自组织神经网络（Self-organization NNs）等

## 4.2 人工神经网络模型原理

### 4.2.1 人工神经元模型

ANN 是应用物理器件外加数学关系模仿生物神经网络结构，从而达到预期

功能的一种模型。神经元是 ANN 中的一个最基本的处理单元，其结构示意图如图 4-1 所示：



4-1 人工神经元模型

上述 M—P 模型是 1943 年神经生物学家、心理学家 W.S.Mcculloch 和数理逻辑学家 W.Pitts，在分析总结神经元基本特性的基础上首先提出的，它是目前大多数常见的神经网络模型的基础。

$x_i$ ——代表神经元  $j$  的输出，也是神经元  $i$  的一个输入；

$w_{ij}$ ——代表神经元  $i$  与神经元  $j$  之间的连接权值(类似与生物神经元之间突触连接的强度)；

$u_i$ ——代表神经元  $i$  的活跃值，即神经元状态；

$\theta_i$ ——代表神经元的阈值；

$\varphi(\bullet)$ ——为激活函数；

$y_i$ ——是神经元  $i$  的输出；

ANN 的数学式表达为： $u_i = \sum_{j=1}^n w_{ij}x_j$ ； $v_i = u_i + \theta_i$ ； $y_i = \varphi(w_{ij}x_j + \theta_i)$

(其中  $v_i$  为神经网络偏差调整后的值)

神经元在获得输入以后，它应当给出适当的输出，在神经网络中的显示过程为：当输入的信号大于神经元本身的阈值时，神经元处于激活状态，否则处于抑制状态，为了有更为宽泛的使用层面，使用更为一般的变换函数来执行神经元的

网络输出，这就是激活函数的原理。一般来说不同的神经网络结构有着不同激活函数，典型的激活函数有以下三种：

1) 阈值函数 (Threshold Function)

$$\varphi(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad \text{常称使用阈值函数神经网络模型为 M-P 模型。}$$

2) 线性函数 (Linear Function)

$$\varphi(x) = \begin{cases} 1 & x \geq 1 \\ A(1+x) & -1 < x < 1 \\ 0 & x \leq -1 \end{cases}$$

线性函数类似于一个带幅度为 A 的线性放大器，使得神经元获得适当的放大作用。

3) S 型函数 (Sigmoid Function)

最常用的 S 型函数形式为： $\varphi(x) = \frac{1}{1 + \exp(-\alpha x)}$ ；另一种为双曲正切函数：

$$\varphi(x) = \frac{1 - \exp(-\alpha x)}{1 + \exp(-\alpha x)}$$

其中  $\alpha$  表示调整曲线斜率；S 型函数具有保持函数单调性的

同时使其具有平滑和渐进性等特性，常用于非线性关系映射中。

#### 4.2.2 神经网络分类

常用的 ANN 从结构方式上可以分为以下几种<sup>[47]</sup>：

①向前神经网络：这种网络结构一般分为三层，输入层，隐含层和输出层，但是隐含层可能是多层的。网络信号是从前向后一层层传递的，最终输出层得到结果。

②有反馈的前馈式神经网络：这种网络结构是分层的，最明显的特征是网络中存在一个从输出层到输入层的一条回路，而且每层的神经元只与下一层的神经元相连，同一层神经元之间没有联系。

③前馈式内层互连神经网络：这种神经网络与其他神经网络最大的不同是在该网络结构中，同一层神经元之间有着相互联系、相互影响，但从整个结构上来看它的主要联系也是层与层之间的连接，大多 Self-organization NNs 都是采用这种结构。

④反馈式全互连神经网络:这种网络结构是说在整个网络中,每个神经元都与其他神经元相连,这样看起来就没有了层的界限,从而使之形成了连续反馈关系,这种结构的网络一般都具有很强的自动寻优功能。

从目前的应用情况来看,有反馈的前馈式神经网络不光是最早被人们认识,而且的应用最为广泛的一种网络,它体现了神经网络最为精华的部分。1986年,Rumelhart 和 McClelland 建立了并行式分布处理理论,提出了误差反向传播神经网络(简称 BP 神经网络),这种网络是根据非线性激活函数通过输出层的误差反向传播反馈使得神经元之间的权值进行调整得出的。在 ANN 的实际应用中,75%-85%的 ANN 模型是采用的 BP 网络模型,这也使得 BP 算法成为了当前神经网络领域最为主要的研究方向。

#### 4.2.3 神经网络的学习规则

ANN 最大的优点就是有自学能力。网络通过对样本数据的训练不断调整神经元之间的权值,这里按照什么方式调整就是我们所说的学习规则。网络训练的过程,就是神经元学习的过程。学习规则就是让神经元之间的关系按照一定的规则进行变动,直至达到最后的一定目标。ANN 是通过模拟人脑处理事物的能力,通过改善自身的性能参数和自身与各个神经元之间的权值最终达到优化网络的目的。这需要通过训练大量的数据样本来实现,训练好的网络中会储存相关有效的网络参数。ANN 的学习方式主要有三种:有监督学习、无监督学习、以及强制学习。ANN 的学习规则一般有两种经典的模式:Hebb 学习规则和 Delta 学习规则。

##### 1) Hebb 学习规则

Hebb 学习规则的基本思想是:两个有联系的神经元之间权值的变化是根据两个神经元是否被同时激活,如果神经元同时激活则权值的变化与激励的强度成一定的比例。其数学描述为:  $\Delta w_{ij}(k) = \eta x_i(k) y_j(k)$ , 其中  $y_j(k)$ ,  $x_i(k)$  分别为神经元  $i$  和神经元  $j$  的激活值,  $w_{ij}$  为神经元  $i$  和神经元  $j$  的连接权值, 其中  $\eta$  为学习步长。

##### 2) Delta 学习规则

Delta 学习规则也叫做误差修正学习规则,它的学习过程如下所示:不防设

实际输出  $y_i(k)$  为输入是  $x_i(k)$  的神经元  $i$  在  $k$  时刻经过处理得到的,  $d_i(k)$  表示输出的目标值, 则输出误差表示为:  $e_i(k) = d_i(k) - y_i(k)$ 。学习的目的就是要使输出误差的目标函数达到最小值, 也就是说所有的神经元经过处理后的输出值最好能逼近要求输出的目标值, 整个过程其实就可以看成一个最优化问题。很容易通过梯度下降法可得学习规则为:  $\Delta w_{ij}(k) = \eta x_i(k) e_i(k)$ 。

## 4.3 BP 神经网络

### 4.3.1 BP 神经网络结构及工作原理

BP 神经网络也被称为误差反向传播神经网络, 是目前运用最为广泛的一种 ANN, 我们常用的 ANN 模型中有相当大一部分采用了 BP 网络的变化形式。BP 网络的结构是这样的: 它是由一个输入层, 一个或者多个隐含层及一个输出层组成的。每层又由很多神经元构成, 但他们之间没有联系, 每层的神经只与前后层的神经元相互联系, 网络按照有监督的学习模式进行训练。其学习过程分两个步骤完成。

1) 网络信号的正向传播阶段——初始信号经输入层至隐含层最终传递到输出层, 再有输出层传出处理后的信号, 这就是完整信号传播的路径。正向传播阶段的网络中神经元之间的权值是不变化的, 上一层的神经元只能影响下一层的神经元, 下一层神经元并不能影响上一层神经元, 如果最终的输出层信号并不能达到预期设定的目标, 则转入下一阶段。

2) 误差信号的反向传播阶段——与网络信号的正向传播阶段不同的是, 误差信号是由网络末端逐层向前传导的, 这就是误差信号的反向传播阶段的传播路径。在此过程中, 网络中各神经元的权值有误差信号反馈动态的纠正, 直至网络能够达到预期设定的目标为止, 否则再次进入网络信号的正向传播阶段。

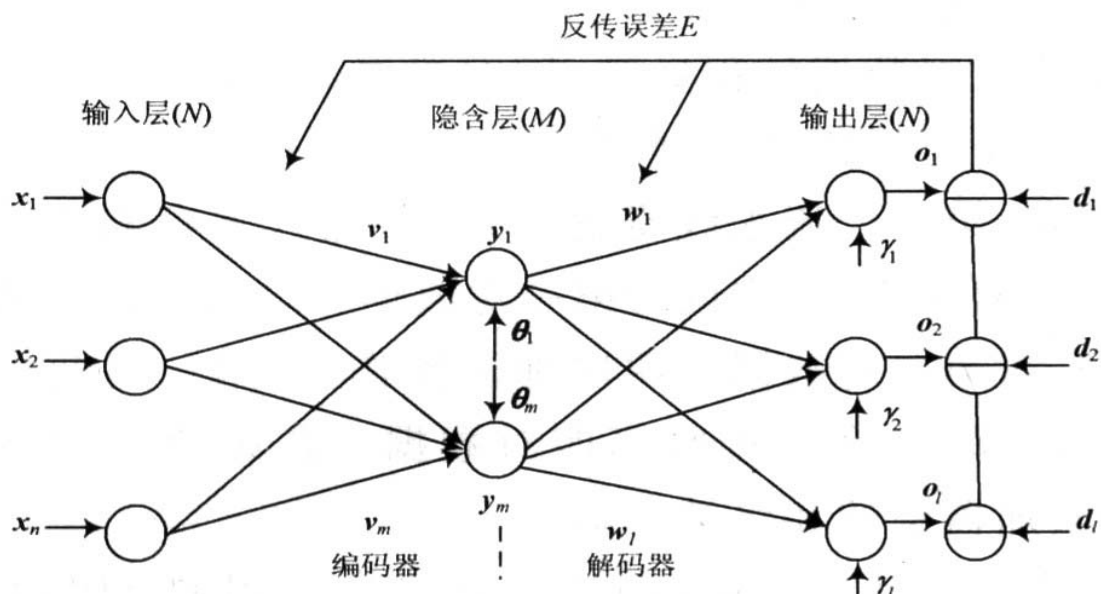


图 4-1 BP 神经网络结构图

### 4.3.2 BP 神经网络学习算法的基本流程

运用 BP 神经网络进行实验时，就必须要先确定神经元相关参数。如网络的结构（有几层隐含层）、网络的激活函数、学习速率等。

以三层全连接 BP 神经网络为例，假设此该网络基本参数如下：

输入层有  $n$  个神经元，

隐含层有  $n_1$  个神经元，

输出层有  $m$  个神经元。

输入层神经元  $i$  的输出值表示为  $Y_i^1$ ；

隐含层神经元  $j$  的输出值表示为  $Y_j^2$ ；

输出层神经元  $k$  的输出值表示为  $Y_k^3$ ；

输出层神经元对应的反馈信号表示为  $T_k$ ；

输入层神经元  $i$  和隐含层神经元  $j$  间的连接权值表示为  $W_{ij}$ ；

隐含层神经元  $j$  和输出神经元  $k$  间的连接权值表示为  $W_{jk}^*$ ；

隐含层神经元  $j$  的阈值表示为  $\theta_j$ ；

输出层神经元  $k$  的阈值表示为  $\theta_k$ 。

神经元的激活函数一般使用 Sigmoid 函数:  $\varphi(x) = \frac{1}{1 + \exp(-\alpha x)}$

误差函数一般选用平方误差函数:  $E = \frac{1}{2} \sum_{k=1}^N \|T_k - Y_k^3\|^2$

### BP 神经网络算法的基本流程

#### ①初始化网络参数

设定初始学习次数  $t=0$ , 网络中各神经元之间的初始权值设为  $(-1,1)$  区间的随机数,  $W_{ij} \in [-1,1]$ ,  $W_{jk} \in [-1,1]$ ,  $\theta_j \in [-1,1]$ ,  $\theta_k \in [-1,1]$ 。

#### ②网络信号的正向传播过程

向已经初始化后的网络中输入一个训练样本  $(X_k, T_k), k \in \{1, 2, \dots, N\}$ , 其中  $N$  为训练样本个数,  $X_k$  为样本输入信号,  $T_k$  为网络输出目标值,  $X_k \in R^n, T_k \in R^m$ 。

隐含层神经元  $j$  的输出值:  $Y_j^2 = f(\sum_{i=1}^n W_{ij} Y_i^1 - \theta_j) = f(\sum_{i=1}^n W_{ij} X_{ki} - \theta_j), j = \{1, 2, \dots, n_1\}$

输出层神经元  $k$  的输出值:  $Y_k^3 = f(\sum_{j=1}^{n_1} W_{jk}^* Y_j^2 - \theta_k), k = \{1, 2, \dots, m\}$

#### ③误差信号的反向传播过程

输出层神经元  $k$  的误差反馈:  $\delta_k = (T_k - Y_k^3) Y_k^3 (1 - Y_k^3), k \in \{1, 2, \dots, m\}$

隐含层神经元  $j$  的误差反馈:  $\delta_j = Y_j^2 (1 - Y_j^2) \sum_{k=1}^m \delta_k W_{jk}^*$

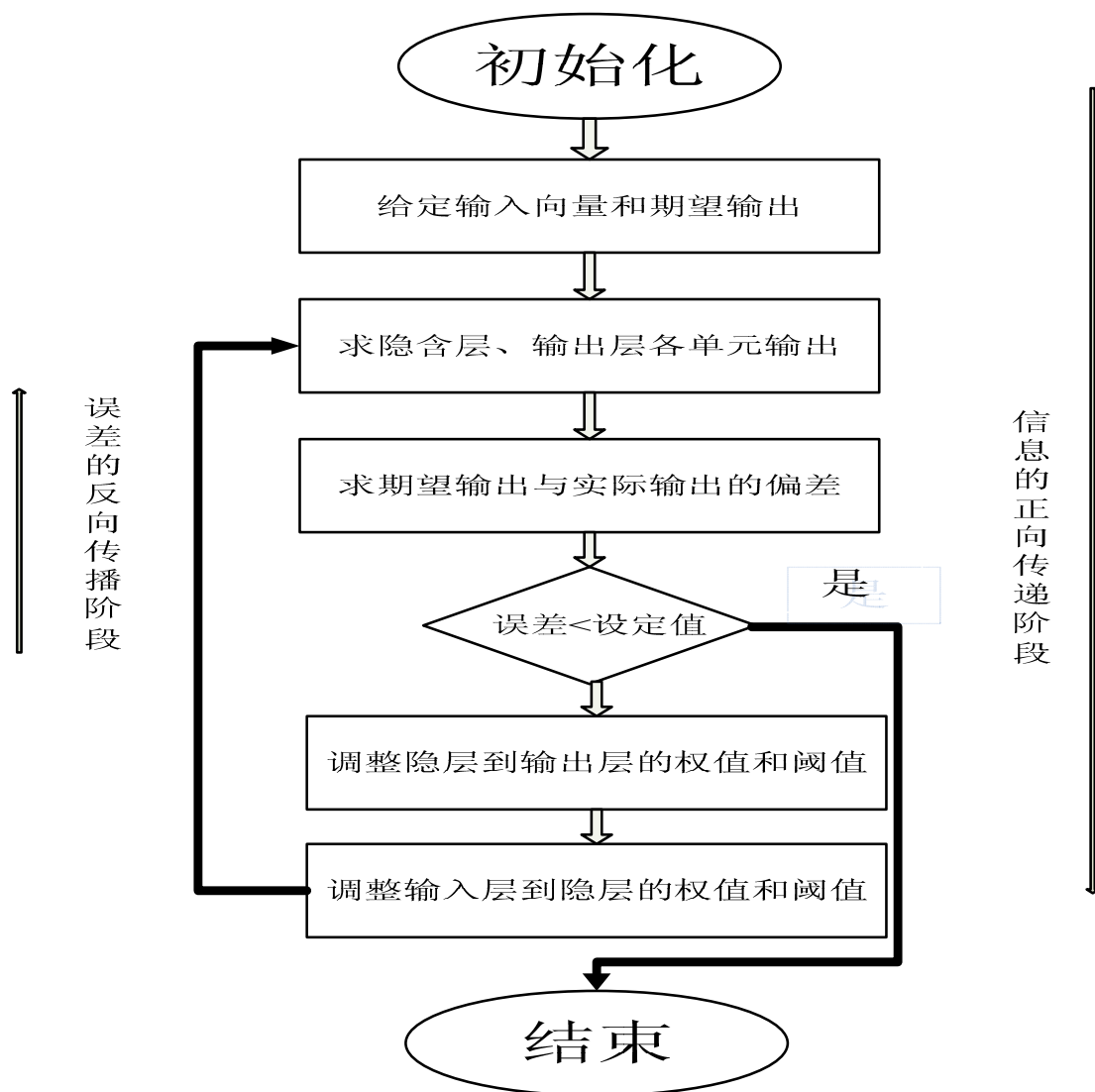


图 4-2 BP 神经网络模拟过程流程图

通过误差信号  $\delta_k$  来修正隐含层和输出层之间的连接权值矩阵  $W_{jk}$  以及神经元

阈值矩阵  $\theta_k$ 。其数学表达为： $W_{jk}^*(t+1) = W_{jk}^*(t) + \alpha \delta_k Y_j^2$ ； $\theta_k(t+1) = \theta_k(t) + \beta \delta_j$

通过误差信号  $\delta_j$  来修正隐含层和输入层的连接权值矩阵  $W_{ij}$  以及神经元阈值矩阵

$\theta_j$ 。其数学表达为： $W_{ij}(t+1) = W_{ij}(t) + \alpha \delta_j Y_i^1$ ； $\theta_j(t+1) = \theta_j(t) + \beta \delta_j$

完成上述所有过程，查看系统的误差  $E = \frac{1}{2} \sum_{k=1}^N \|T_k - Y_k^3\|^2$ ，看系统误差是否小于

所要求的误差最大值，若已经小于规定的误差最大值，则网络训练结束；当然如果网络训练的次数已经达到规定的训练的最大次数，则网络训练结束；否则将网络训练次数更新为  $t = t + 1$ ；重复上述过程。



### 4.3.3 BP 神经网络的缺陷

BP 神经网络有诸多优点,但从上述介绍中我们可以看出其中也有很多缺陷,下面就此简单介绍一下:

第一、BP 神经网络适用于非线性预测模型,这样 BP 神经网络就可能陷入局部极小的困境。BP 神经网络中神经元之间的权值确定是通过梯度下降法求得的,这样就造成了不同的起点会造成不同点最优解。所以说 BP 算法很可能就是寻求的局部最优值而不是整体的最优值。

第二、并不是所有的非线性问题都可以通过 BP 神经网络解决。很多时候因为问题自身的原因可能造成 BP 神经网络的收敛速度很慢甚至不收敛的情况。有一些特殊问题在 BP 神经网络的训练过程中可能会造成激活函数饱和、网络权值调整失效等问题。实验表明 BP 神经网络的收敛速度于网络的初始权值的选取有很大的关系。

第三、网络结构的确定没有理论的指导。初始化网络的时候到底用使用几层隐含层比较合适,每层的神经元个数怎么确定,这些都没有现成的公式,只能通过尝试比较确定。一般情况下,如果网络规模过大会导致网络容错性降低,训练时间过长。如果网络规模过小,甚至可能造成网络不能收敛的情况发生。

第四、对于训练的样本要求比较高。首先要有足够的样本数量使得网络能够完成,其次要求样本有相同的特征数。再者新样本的加入可能会导致网络的重置。

根据以上描述,可以发现 BP 网络主要存在,易陷入局部最小值、收敛速度过慢(网络不收敛)、网络结构无法确定、对于训练样本要求过高等问题。针对上述问题很多学者给出了改进型的 BP 算法。在 MATLAB R2009a 版本的神经网络工具箱中就给出了在中能够直接应用的改进型的 BP 网络。

## 4.4 小波神经网络

### 4.4.1 小波神经网络概述

小波分析(Wavelets Analysis)是上世纪 80 年代发展起来的一门技术,它的发现标志着傅里叶变换进入了一个新的时期。它的具有良好的时频局部特性和变焦特性,是一种多分辨率分析方法。但是因为在现实应用中对小波基的构建和储

存不是很方便，而且变换时抛弃了时间信息，这使得小波理论只能在很窄的领域得到应用。神经网络的诞生就解决了这个问题，神经网络极强的自组织性和自学习性，能够很轻松解决记忆、推理等功能。所以两种理论相互结合的成果小波神经网络就这样孕育而生。

小波神经网络具有小波变换和神经网络的双重优势，它作为一种新型的数学理论和分析方法，在众多领域引起了极大的关注。小波分析可以通过平移和伸缩变换逼近任意函数使其具有了良好的精度和容错性，而且在维空间有选择信号方向的能力。目前两者能通过两种方式相互结合：第一“松散型”结合方式，先对数据信号做小波分析处理，再把处理完的信号输入神经网络中；第二“紧致型”结合方式，用小波基函数代替神经网络隐含层众多激活函数，这种模型同时具有了神经网络和小波变换的优点。因为“紧致型”小波神经网络处理数据的能力更强，所以“紧致型”小波神经网络成为了主要的研究方向。

小波神经网络是以 BP 神经网络为模型基础，把小波基函数作为神经网络隐藏层的激活函数的一种新型神经网络，它与 BP 神经网络最大的不同就是他们的隐藏层的激活函数不一样。这样看来小波神经网络是一种有反馈的前馈式神经网络，小波变换可以使网络具有很快的学习速度同时也避免了数据陷入局部极小的缺陷，使其具有了很广泛的应用领域。小波神经网络的理论依据如下：

在 Hilbert 空间中选取一个母小波函数（基本小波函数） $\psi(t)$ ，使其满足

$$\int_{\mathbb{R}} \frac{|\bar{\psi}(w)|^2}{w} dw < \infty \quad (\bar{\psi}(w) \text{ 为 } \psi(t) \text{ 的 Fourier 变换})。通过 \psi(t) \text{ 的伸缩和平移变换产生小波函数基：}$$

$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right)$ ；其中  $a$ 、 $b$  分别为伸缩尺度因子和平移因子。

$$\text{小波神经网络激活函数的数学描述：} y_k = \sum_{j=1}^J w_{kj}^* \psi \left( \frac{\sum_{i=1}^I w_{ji} x_i - b_j}{a_j} \right) \quad 4-1$$

其中， $x_i (i=1,2,\dots,I)$  为输入层第  $i$  个神经元输入信号；

$y_k (k=1,2,\dots,k)$  为输出层第  $k$  个神经元的输出信号；

$w_{ji}$  为隐含层神经元  $j$  和输出层神经元  $i$  之间的权值;

$w_{kj}^*$  为输入层神经元  $j$  和隐含层结点  $k$  之间的权值;

$a_j$  为第  $j$  个隐含层神经元的伸缩尺度因子;

$b_j$  为第  $j$  个隐含层神经元的平移因子;

#### 4.4.2 小波神经网络算法的基本流程

设  $P(p=1,2,\dots,n)$  为输入信号样本的个数,  $x_i^p$  为第  $p$  个样本的输入值,  $y_k^p$  为第  $p$  个样本的网络输出值,  $t_k^p$  为第  $p$  个样本的输出目标值。通过网络训练确定网

络参数  $w_{ji}$ 、 $w_{kj}^*$ 、 $a_j$  和  $b_j$ , 使得误差  $E_k^p = \frac{1}{2} \sum_{k=1}^K (t_k^p - y_k^p)^2$  可以达到最优值。

小波神经网络算法的基本流程<sup>[48]</sup>:

##### ①初始化网络参数

设定初始学习次数  $t=0$ , 输入层与隐含层权值  $w_{ji}$ 、隐含层与输出层的权值  $w_{kj}^*$ , 伸缩尺度因子  $a_j$  和平移因子  $b_j$  的初始值设为  $(-1,1)$  区间的随机数,  $w_{ji} \in [-1,1]$ ,  $w_{jk}^* \in [-1,1]$ ,  $a_j \in [-1,1]$ ,  $b_j \in [-1,1]$ 。

##### ②预测网络输出、计算预测误差

向已经初始化后的网络中输入一个训练样本  $(P_k, T_k), k \in \{1, 2, \dots, N\}$ , 其中  $N$  为训练样本个数,  $P_k$  为样本输入信号,  $T_k$  为网络输出目标值,  $P_k \in R^n, T_k \in R^m$ 。

根据式 4-1 计算网络输出; 根据  $E_k^p = \frac{1}{2} \sum_{k=1}^K (t_k^p - y_k^p)^2$  计算预测误差

##### ③网络权值修正

根据  $E_k^p$  的情况利用最速下降法究竟小波神经网络的参数  $w_{ji}$ 、 $w_{kj}^*$ 、 $a_j$  和  $b_j$  的变化情况:

$$\text{令 } x = \frac{\sum_{i=1}^I w_{ji} x_i^p - b_j}{a_j}$$

$$\nabla w_{kj}^* = \frac{\partial E_k^p}{\partial w_{kj}^*} = -(t_k^p - y_k^p) \psi(x)$$

$$\nabla w_{ji} = \frac{\partial E_k^p}{\partial w_{ji}} = -\sum_{k=1}^K (t_k^p - y_k^p) w_{ji} \frac{\partial \psi(x)}{\partial x} \frac{x_i^p}{a_j}$$

$$\nabla b_j = \frac{\partial E_k^p}{\partial b_j} = \sum_{k=1}^K (t_k^p - y_k^p) w_{kj} \frac{\partial \psi(x)}{\partial x} \frac{1}{a_j}$$

$$\nabla a_j = \frac{\partial E_k^p}{\partial a_j} = \sum_{k=1}^K (t_k^p - y_k^p) w_{kj} \frac{\partial \psi(x)}{\partial x} \frac{x_i^p}{a_j^2}$$

为了加快网络的收敛速度，引入一个网络参数动量因子  $\alpha$ ，因此权向量的迭代公式为：

$$w_{kj}^*(t+1) = w_{kj}^*(t) - \eta \nabla w_{kj}^* + \alpha \Delta w_{kj}^*$$

$$w_{ji}(t+1) = w_{ji}(t) - \eta \nabla w_{ji} + \alpha \Delta w_{ji}$$

$$b_j(t+1) = b_j(t) - \eta \nabla b_j + \alpha \Delta b_j$$

$$a_j(t+1) = a_j(t) - \eta \nabla a_j + \alpha \Delta a_j$$

完成上述所有过程，查看系统的误差  $E_k^p = \frac{1}{2} \sum_{k=1}^K (t_k^p - y_k^p)^2$ ，看系统误差是否小于所要求的

误差最大值，若已经小于规定的误差最大值，则网络训练结束；当然如果网络训练的次数已经达到规定的训练的最大次数，则网络训练结束；否则将网络训练次数更新为  $t = t+1$ ；重复上述过程。

## 4.5 BP 神经网络和小波神经网络模型的应用

本章使用的原始数据选取了从1997-2010年的青岛市房地产市场开发数据作为研究依据，2011-2012年的数据作为检验样本数据序列，预测2013-2016年的房地产市场相关参数指标。本节以房屋施工面积为预测对象进行说明，房地产开发投资额、房屋实际销售面积这两个指标的计算过程类似。

### 4.5.1 BP 神经网络的应用

青岛市的房屋施工面积的历史数据必然与未来的房屋施工面积有着某种关系，用数学方法来表示前  $k$  年房屋施工面积  $(x_{n-k}, x_{n-k-1}, \dots, x_{n-1}, x_n)$  与第  $n+t$  年的

房屋施工面积  $x_{n+t}$ ，之间存在的函数关系表示为： $x_{n+t} = \phi(x_{n-k}, x_{n-k-1}, \dots, x_{n-1}, x_n)$ 。这种关系很明显可以通过 ANN 来模拟，本节 BP 神经网络是基于时间序列趋势的一种预测模型。

本文中采用的样本是 1997——2010 年的青岛市房屋施工面积数据，通过训练好的 BP 网络对 2011——2012 年的房屋施工面积进行预测，进行网络的预测精度检验，最终用该网络预测 2013-2016 年的青岛市房屋施工面积。初始设置网络输出层的神经元个数为五个，输出层有一个神经元，从 1997 年的青岛市房屋施工面积开始每五年的数据作为一个输入样本，依次排列，而第六年的数据作为样本输出目标值，进行网络训练。根据现行的神经网络理论无法明确 BP 神经网络中到底应该有几层的隐含层和每个隐含层中含有多少神经元个数，但有研究证明在隐含层神经元足够的情况下，三层 BP 神经网络拟合任何非线性函数的能力。所以本文中采用三层 BP 神经网络模型，隐含层的神经元个数采用经验公式与试算法进行确定： $N = \sqrt{n+m} + a$  其中  $N$  为隐含层的神经元个数， $n$  为输入层神经元个数， $m$  为输出层神经元个数，通常取  $0 < a < 10$ 。通过试算法本文隐层节点数确定为 9，设定的相关网络参数为：其中学习速率  $lr=0.05$ ，动量系数  $mc=0.9$ ，最大学习次数为 10000，学习误差  $E = 10^{-8}$ 。

因为网络训练的样本物理量不尽相同，其数值大小级数相差很大，使得各指标没有可比性。另外未来提高网络的性能和加快网络的训练速度，一般都对样本数据进行归一化处理<sup>[49]</sup>。一般是把输入样本和网络输出值都归一化到  $[-1,1]$  区间内。这样既保持着样本数据的特征，又使得样本更适合于神经网络的使用条件。当前的 Matlab 软件中集成了归一化函数：

`[inputn,inputps]=mapminmax(input_train)。`

从表 4-1 可以看出 BP 网络模型对的波动性较大数据预测效果比较好，但仍然存在 5%左右的误差。这是因为 ANN 只是将预测的系统结构看作一个黑箱结构，在预测过程中只是考虑了输入和输出的关系（趋势性的关系），而没有考虑其影响因素和系统的确定性信息。这样看来造成一定的系统误差也在情理之中，如果考虑其影响因素是不是可能会取得较好的效果呢？下面就小波神经网络进行因素定量分析的方式进行预测。

表 4-1 2011—2012 年青岛市房屋施工面积

年份	2011 年			2012 年		
	实际数据	BP 模型 预测值	相对误差 百分比	实际数 据	BP 模型 预测值	相对误差 百分比
房屋施工面积	5693	5634	-1.04%	6473	6412	-0.94%
房地产开发投资 额	782.72	839.20	7.22%	930.10	960.08	3.22%
房屋实际销售面 积	1026	1073.54	4.63%	950.9	886.23	-6.80%

根据训练好的 BP 网络预测 2013-2016 年的房屋施工面积数据如下表所示：

表 4-2 2013-2016 年青岛市房屋施工面积

年份	2013 年	2014 年	2015 年	2016 年
房屋施工面积	7513	8256	9685	9816
房地产开发投资额	1055	1227	1415	1715
房屋实际销售面积	1174	960	1210	1053

#### 4.5.2 小波神经网络的应用

从上节的结论中我们可以看出，如果仅通过 BP 神经网络对时间序列进行趋势性预测得到的结果可能并不是很理想，下面这节就上述问题用小波神经网络通过因素定量的方式来进行预测。根据青岛市房地产市场发展状况，结合文章研究内容所需数据的可得性、易于量化方面考虑，确定了 16 组房地产市场的影响要素分别为：人口密度、已婚有生育能力的人数、社会从业人数、普通高等学校在校人数、人均 GDP、平均工资、进出口总额、地方财政收入、金融系统人民币贷款总额、居民消费和零售价格指数、消费品零售总额、规模以上固定资产投资总额、建筑安装工程投资总额、年末道路面积、全年用电量、全年客运量。

本文中采用的样本是 1997—2010 年的上述 16 组数据，通过训练好的小波神经网络对 2011—2012 年的房屋施工面积进行预测，并进行网络的预测精度检验，最终用该网络预测 2013-2016 年的青岛市房屋施工面积。初始设置网络输出层的神经元个数为 16 个，输出层有 3 个神经元：分别是房屋施工面积、房地产开发投资额、房屋实际销售面积。设置隐含层为 28 个神经元，即有 28 个小波基函数，设定的相关网络参数为：其中学习速率  $lr=0.05$ ，动量系数  $mc=0.9$ ，最大学习次数为 10000，学习误差  $E=10^{-8}$ 。由于 Morlet 小波基函数函数对称且表达式简单，具有计算稳定和对干扰的鲁棒性好等优点，故本实验采用的激活

函数为 Morlet 母小波，即  $\varphi = \cos(1.75x)e^{-\frac{x^2}{2}}$ 。

选取 1997~2010 年的样本数据，通过数据归一化，然后经过网络训练，还原预测数据得到 2011~2012 年青岛市房屋施工面积数据。

表 4-3 2011~2012 年青岛市房屋施工面积

年份	2011			2012		
	实际数据	小波神经网络预测值	相对误差百分比	实际数据	小波神经网络预测值	相对误差百分比
房屋施工面积	5693	5658	-0.61%	6473	6464	-0.14%
房地产开发投资额	782.72	802	2.46%	930.10	953.00	2.46%
房屋实际销售面积	1026	998	-2.73%	950.9	973.00	2.32%

从表 4-1 和表 4-3 的对比中可以看出小波神经网络在实验精度上明显比 BP 神经网络好。而且从图 4-3 小波神经网络的运行状态图中可以看出，要达到相同目标误差（0.00001），小波网络的训练次数为 29 次，BP 神经网络则需要 100 多次次，这表明在收敛速度上小波神经网络也具有明显的优势。从实验的整体情况来看，因为 BP 神经网络和小波神经网络拥有着类似的网络结构，但是因为激活函数的不同使得小波神经网络的预测效果明显好于 BP 神经网络，另外对于像房屋建筑面积这种比较复杂的系统指标，受到外界干扰因素还是比较大的，从实验的结果表明因素定量分析预测优于趋势分析预测。

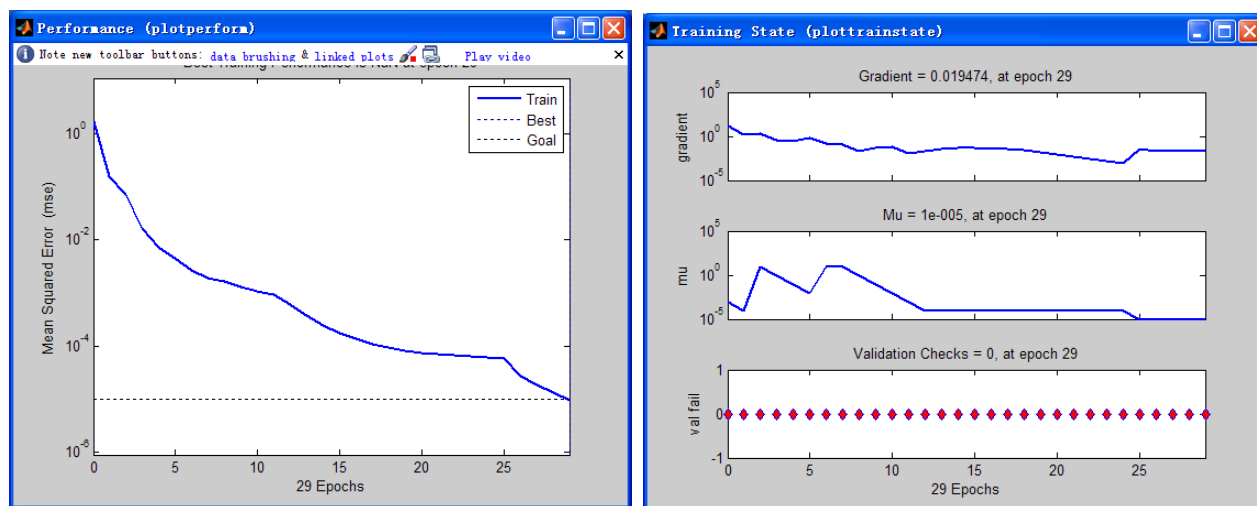
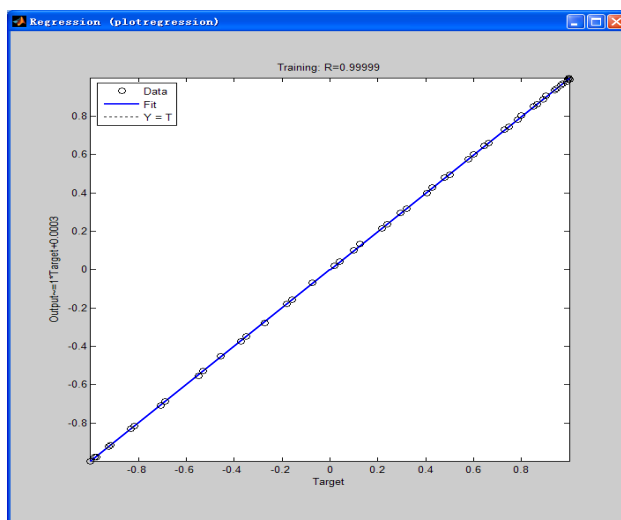


图 4-3 小波神经网络的运行状态



对于房地产市场的影响要素一般随着时间变化呈现出某种特定的规律,在本次实验中运用了一次线性回归和灰色预测模型对 2013-2016 的 16 组房地产影响要素进行初步的判定。其中人均 GDP、平均工资、进出口总额、地方财政收入、金融系统人民币贷款总额、居民消费和零售价格指数、消费品零售总额、规模以上固定资产投资总额、建筑安装工程投资总额、年末道路面积、全年用电量、全年客运量投资采用灰色预测模型,其他影响因素均采用一元线性模型。

将所需预测的样本输入已经训练好的小波神经网络,可以预测 2013-2016 年青岛市房屋施工面积预测值如表 4-4 所示:

表 4-4 2013-2016 年青岛市房屋施工面积预测值

年份	2013 年	2014 年	2015 年	2016 年
房屋施工面积	7352	8423	9521	10051
房地产开发投资额	1125	1352	1435	1653
房屋实际销售面积	1064	988	1187	1120

## 4.6 本章小结

本章通过对神经网络基础概念的介绍,对其结构原理、学习规则有了一定的理解。在了解神经网络基本构造的基础上,介绍了 BP 神经网络和小波神经网络的工作原理和算法流程,简要的说明了两者的区别和联系。最后运用这两种神经网络分别对青岛市房地产市场进行相关预测,为下一章组合预测提供了合格的数据样本。



## 5 组合预测模型

### 5.1 组合预测的概念及意义

面对不同的问题，一般根据问题本身的特性应采取不同的预测模型进行处理，其预测效果往往也是有所差异的。没有一种预测模型可以解决所有问题，甚至对于同一个问题用同一种预测方式在不同点时刻预测出来的效果也不尽相同。这不光是因为现实中的问题本来就是多样的，而且是由于各种预测模型的建模机理也是不同的。每一种预测模型只是按照其预设的原理来对现实的现象进行推导，再此过程中肯定会有很多重要的信息被遗漏。它们之间是相互联系，互相补充的。自然而然的我们可以想到是不是可以把各种模型进行组合，找到一种集几种单一模型优点的模型。这种预测模型是不是就遗漏了少量的信息，更加全面的反映了问题本身的规律和未来的发展趋势。组合预测的精髓<sup>[50]</sup>：对于同一问题，通过某种方法把各个单一预测模型的信息组合起来，使得能更加合理全面的反映这个问题的内在规律，以期可以提高预测精度。

一般来说组合预测能更加全面的反映一个问题的整体信息，与单一预测模型相比，有更好的稳定性和精度。所谓的组合预测模型就是利用各个单项预测所提供的信息，通过加权平均的方式进行组合，形成一种新的预测方法。只要组合方式恰当，可能就会得到意想不到的结果。在得到各种单项预测模型的预测值后，如何求其加权平均系数就显得格外重要。一般来说，按照加权系数计算方法的不同可将组合预测分为两类：最优组合预测方法和非最优组合预测方法。最优组合预测的基本思想：通过类似线性规划的方式，构造能够使预测结果最好的组合参数，从而求得最佳加权系数。但是这种方法有一种很明显不合理的地方，就是加权系数可能是负值。这样的话被赋予负值的单一模型就没有任何的实际意义，仅仅是为了“拼凑”最佳模型的精度而已。非最优组合预测的基本原理<sup>[51]</sup>：单项预测模型误差的方差与其权系数成反比。一般来说非最优组合预测精度要比最优组合预测差，但是非最优组合预测在原理上更容易被人接受。

组合预测的基本思想：

记  $y_{it} (i=1,2,\dots,m; t=1,2,\dots,n)$  表示第  $i$  种预测方法在  $t$  时刻得到的预测值，则  $m$  种

预测方法的组合预测值为： $\hat{y}_t = \sum_{i=1}^m W_i \hat{y}_{it} (t=1, 2, \dots, n)$ ，其中  $\sum_{i=1}^m W_i = 1$  其中  $W_i$  是第  $i$  种预测方法的权系数。

通过上述分析中可以看出，组合预测研究的最主要的内容就是各单项预测方法的权重的确定，权重的大小表示单项预测在组合预测中的重要性。一般来说组合预测是在单项预测不能得到合理预测结果或者不能肯定几种单项预测方法效果的情况下使用的。在大多数情况下，其实组合预测模型并不比其被组合的某种精度最高的单项预测模型预测效果好，只是组合预测更全面从多个角度分析了问题的发展趋势，在数值上的体现可能并不是最优的<sup>[52]</sup>。在单项预测模型不能确定其预测精度的情况下，通过组合预测也不愧为一种有效解决问题的有效途径。当然，如果能找到一个预测模型很好的反映所需预测问题的内在规律和发展趋势，就没有必要在使用组合预测画蛇添足了。相反如果组合预测中有一项预测误差很大，要对其舍弃，不然会对整个预测效果造成很大的干扰。

本文利用信息论中信息熵的原理计算出组合预测模型的加权平均系数，它属于一种非最优正权组合预测方法。

## 5.2 熵值法组合预测模型

组合预测核心的问题就是加权平均系数如何得到，这关系着组合预测模型的结构合理程度和结果的精确程度。非最优组合预测的基本思想就是单项预测模型误差序列变异程度越大，其在组合预测中的权重就越小。一般来说组合预测模型的加权平均系数直接与单项预测所包含的信息直接相关，但是这又如何度量呢？在通常情况下变异程度用残差或方差等统计指标来进行度量，本节利用信息论中的熵值概念反映其变异程度的一个指标。在信息论中，熵值反映了系统状态的无序程度、不确定程度和信息缺乏程度的一个状态量，即体现了系统某个指标的变异程度<sup>[53]</sup>。熵值越小，其包含的信息就越大，在组合预测中的权重就越大，反之就权重越大。如此说来把熵值法应用到房地产市场预测中，可能会有意想不到的效果。

熵值法确定组合预测加权平均系数步骤<sup>[54]</sup>：

设对同一预测目标的某个指标序列  $\{x_t, t=1, 2, \dots, N\}$ ，通过  $m$  种单项预测模型对其进行预测。

设第  $i$  种单项预测模型在  $t$  时刻的预测值为  $x_{it}, i=1, 2, \dots, m, t=1, 2, \dots, N$

$$\text{第 } i \text{ 种预测模型第 } t \text{ 时刻的相对误差 } e_{it} = \begin{cases} 1, & \left| \frac{x_t - x_{it}}{x_t} \right| \geq 1 \\ \left| \frac{x_t - x_{it}}{x_t} \right|, & 0 \leq \left| \frac{x_t - x_{it}}{x_t} \right| < 1 \end{cases}$$

显然  $e_{it}$  大于 0 小于 1,  $\{e_{it}, t=1, 2, \dots, N\}$  为第  $i$  种预测模型在  $t$  时刻的预测相对序列,  $i=1, 2, \dots, m$

①单项预测模型的预测相对误差序列归一化处理。

计算第  $i$  种单项预测模型在第  $t$  时刻的预测的相对误差比重。

$$p_{it} = \frac{e_{it}}{\sum_{t=1}^N e_{it}}, t=1, 2, \dots, N, \text{ 显然, } \sum_{t=1}^N p_{it} = 1, i=1, 2, \dots, m$$

②单项预测模型的预测相对误差的熵值计算。

$$h_i = -k \sum_{t=1}^N p_{it} \ln p_{it}, i=1, 2, \dots, m, \text{ 其中 } k > 0 \text{ 为常数, } h_i \geq 0, i=1, 2, \dots, m.$$

对第  $i$  种单项预测模型来说, 如果其预测的相对误差比重  $p_{it}$  全部相等, 那么熵值  $h_i$  取极大值, 即  $h_i = k \ln N$ , 一般这里取  $k = 1/N$ , 则有  $1 \geq h_i \geq 0$ 。

③单项预测模型的预测相对误差序列的变异程度系数计算。

由于预测对象的熵值与其所含信息量呈相反相关原因, 所以定义第  $i$  种单项预测模型的预测相对误差序列的变异程度系数  $d_i$  为:  $d_i = 1 - h_i, i=1, 2, \dots, m$ ,

加权平均系数记取原则: 某个单项预测模型预测误差序列的变异程度与其加权平均系数成负相关。

$$\text{计算各个单项预测模型的加权平均系数 } l_1, l_2, \dots, l_m \text{ 为: } l_i = \frac{1}{m-1} \left( 1 - \frac{d_i}{\sum_{i=1}^m d_i} \right)$$

显然权系数满足:  $\sum_{i=1}^m l_i = 1$ 。

计算组合预测值  $\bar{x}_t$  :  $\bar{x}_t = \sum_{i=1}^m l_i x_{it}, t=1,2,\dots,N$

### 5.3 基于组合预测的青岛市房地产市场预测

前两章中分别采用 2 种不同的预测思想单独对青岛市房地产市场相关指标进行了预测,从模型对样本数据模拟后的平均相对误差来看,其中从预测效果来看神经网络方法比灰色预测效果好。为了将各单个模型的优势综合起来,使建立的模型更加合乎实际,模型的解释力度更强,本文采用熵值法对改进  $GM(1,1)$ ——马尔柯夫模型、BP 神经网络、小波神经网络的模型进行组合,形成一个组合预测模型。

#### 5.3.1 组合模型的建立

组合函数的确定,关键是各单个预测模型权系数的确定,在此假设  $l_1, l_2, l_3$  别为改进  $GM(1,1)$ ——马尔柯夫模型、BP 神经网络、小波神经网络的模型的权系数,采用熵值法建立线性规划模型如下  $f = l_1 f_1 + l_2 f_2 + l_3 f_3$

本文以房屋施工面积为预测对象进行说明,房地产开发投资额、房屋实际销售面积这两个指标的计算过程类似。

表 5-1 1997-2012 青岛房屋施工面积

年份	房屋施工面积	改进灰色——马尔柯夫		BP 神经网络		小波神经网络	
		预测值	相对误差	预测值	相对误差	预测值	相对误差
1997	675	675	0.00%			671.69	-0.49%
1998	742	872.27	17.56%			738.81	-0.43%
1999	900	889.64	-1.15%			901.31	0.15%
2000	1047	1055.85	0.85%			1046.95	0.00%
2001	1374	1248.11	-9.16%			1370.75	-0.24%
2002	1417	1470.09	3.75%			1416.06	-0.07%
2003	1765	1725.94	-2.21%	1781	0.91%	1771	0.34%
2004	2101	2020.4	-3.84%	2121	0.95%	2110	0.43%
2005	2363	2358.8	-0.18%	2381	0.76%	2358	-0.21%
2006	2751	2747.21	-0.14%	2801	1.82%	2772	0.76%
2007	3224	3192.51	-0.98%	3204	-0.62%	3241	0.53%
2008	3773	3702.48	-1.87%	3741	-0.85%	3766	-0.19%
2009	4310	4285.96	-0.56%	4381	1.65%	4352	0.97%
2010	5058	4952.96	-2.08%	5088	0.59%	5101	0.85%
2011	5693	5602.29	-1.59%	5634	-1.04%	5658	-0.61%
2012	6473	6471.79	-0.02%	6412	-0.94%	6464	-0.14%

注：其中由于 1997-2002 年为 BP 神经网络的训练样本没法得到相关预测值，因此用 2003-2012 年的数据进行模型计算。

由式(4.35)计算三种单一预测方法的预测相对误差的熵值  $h_1$ 、 $h_2$ 、 $h_3$ ：

$$h_1 = -k \sum_{t=1}^{10} p_{1t} \ln p_{1t} = 0.8275, \quad h_2 = -k \sum_{t=1}^{10} p_{2t} \ln p_{2t} = 0.9711, \quad h_3 = -k \sum_{t=1}^{10} p_{3t} \ln p_{3t} = 0.9307$$

由式(4.36)计算三种单一预测方法的预测相对误差序列的变异程度系数  $d_1$ 、 $d_2$ 、 $d_3$ ： $d_1=1-h_1=0.1725$ ， $d_2=1-h_2=0.0289$ ， $d_3=1-h_3=0.0693$ ，

由式(4.37)计算三种单一预测方法的加权系数  $l_1, l_2, l_3$

$$l_1=0.1814, \quad l_2=0.446, \quad l_3=0.372$$

由式(4.38)得组合预测模型为

$$f_{(1)} = 0.1814f_1 + 0.4466f_2 + 0.372f_3 \quad 5-1$$

### 5.3.2 组合预测及结果分析

利用式 5-1 的组合预测模型预测出 2003-2012 年青岛市房屋施工面积，预测结果见表 5-2。

表 5-2 2003-2012 年青岛市房屋施工面积

年份	房屋施工面积	灰色模型	BP 神经网络	小波神经网络	熵值法组合预测
		相对误差	相对误差	相对误差	相对误差
2003	1765	-2.21%	0.91%	0.34%	0.13%
2004	2101	-3.84%	0.95%	0.43%	-0.11%
2005	2363	-0.18%	0.76%	-0.21%	0.23%
2006	2751	-0.14%	1.82%	0.76%	1.07%
2007	3224	-0.98%	-0.62%	0.53%	-0.26%
2008	3773	-1.87%	-0.85%	-0.19%	-0.79%
2009	4310	-0.56%	1.65%	0.97%	1.00%
2010	5058	-2.08%	0.59%	0.85%	0.20%
2011	5693	-1.59%	-1.04%	-0.61%	-0.98%
2012	6473	-0.02%	-0.94%	-0.14%	-0.48%
误差累积		13.46%	10.13%	5.03%	5.24%
平均误差		1.35%	1.01%	0.50%	0.52%

从表 5-2 可以看出，通过组合预测模型预测出来的房屋施工面积平均绝对误差指标值均小于通过灰色预测模型和 BP 神经网络模型单独预测出来该项指标值，但是大于小波神经网络模型预测出来的该项指标值。由于熵值法属于非最优正权的方法，所以这种组合预测模型拟合的结果在数值上看起来不是最优的，但是它综合了三种模型的优点，通过趋势预测和因素预测的结合，在结构和结果上都是合理的。因此，虽然该模型在数值上没有达到目标函数最优化的结果，但是建模的机理和预测的效果都达到了预设目标的，所以该模型预测青岛市房屋施工面积这个指标是非常适合的。

从上表中可以发现，使用 2011-2012 年的外推数据对模型进行检验，发现组合预测模型预测的平均相对误差为 0.73%。模型精度已经很高，因此，该模型可以用来预测青岛市房屋施工面积。

按照上述步骤分别对房地产开发投资额、房屋实际销售面积这两个指标进行预测可得到结果如表 5-3 和 5-4 所示

表 5-3 2003-2012 年青岛房地产开发投资额表

年份	房地产开发投资额	灰色模型	BP 神经网络	小波神经网络	熵值法组合预测
		相对误差	相对误差	相对误差	相对误差
2003	127.80	6.15%	-1.41%	2.51%	3.15%
2004	162.70	3.18%	5.10%	1.42%	2.85%
2005	223.84	-7.34%	1.41%	1.41%	-2.02%
2006	268.36	-4.65%	-0.51%	2.10%	-1.08%
2007	322.35	-2.20%	2.99%	0.82%	0.07%
2008	380.57	1.95%	2.22%	1.17%	1.68%
2009	459.48	3.81%	-0.54%	0.55%	1.61%
2010	602.44	-2.76%	1.42%	-1.90%	-1.56%
2011	782.72	-8.15%	7.22%	2.46%	-0.74%
2012	930.10	-5.21%	3.22%	2.46%	-0.39%
误差累积		45.40%	26.04%	16.79%	15.16%
平均误差		4.54%	2.60%	1.68%	1.52%

房地产开发投资额的组合预测模型为： $f_{(2)} = 0.3923f_1 + 0.219f_2 + 0.4059f_3$

表 5-4 2003-2012 年青岛房屋实际销售面积表

年份	房屋实际销售面积	灰色模型	BP 神经网络	小波神经网络	熵值法组合预测
		相对误差	相对误差	相对误差	相对误差
2003	469	<b>-0.67%</b>	-1.92%	-0.85%	-1.34%
2004	516	<b>5.18%</b>	-0.97%	-0.19%	-0.55%
2005	740	<b>-14.84%</b>	1.08%	0.95%	1.01%
2006	719	<b>1.47%</b>	0.97%	0.83%	0.90%
2007	833	<b>1.11%</b>	0.24%	-0.24%	-0.02%
2008	770	<b>25.95%</b>	-1.04%	-2.47%	-1.82%
2009	1262	<b>-11.72%</b>	-4.83%	-0.87%	-2.68%
2010	1360	<b>-6.09%</b>	-4.56%	-0.59%	-2.40%
2011	1026	<b>42.42%</b>	4.63%	-2.73%	0.62%
2012	950.9	<b>75.49%</b>	-6.80%	2.32%	-1.83%
误差累积			27.05%	12.05%	13.15%
平均误差			2.70%	1.20%	1.32%

注：灰色模型没有通过模型检验所以没有加入组合预测模型中

房地产开发投资额的组合预测模型为： $f_{(3)} = 0.4553f_2 + 0.5447f_3$

可以看出上述模型是可以运用与青岛房地产市场的相关指标的预测的。因此应用此模型对 2013-2016 年间相关数据进行组合预测说得结果如下所示：

表 5-5 2013-2016 年青岛房地产市场发展状况

年份	2013 年	2014 年	2015 年	2016 年
房屋施工面积 (万平方米)	7471.80	8399.83	9680.06	10201.68
房地产开发投资额 (亿元)	1093.36	1315.43	1503.33	1794.30
房屋实际销售面积 (万平方米)	1114.08	975.25	1197.47	1089.49

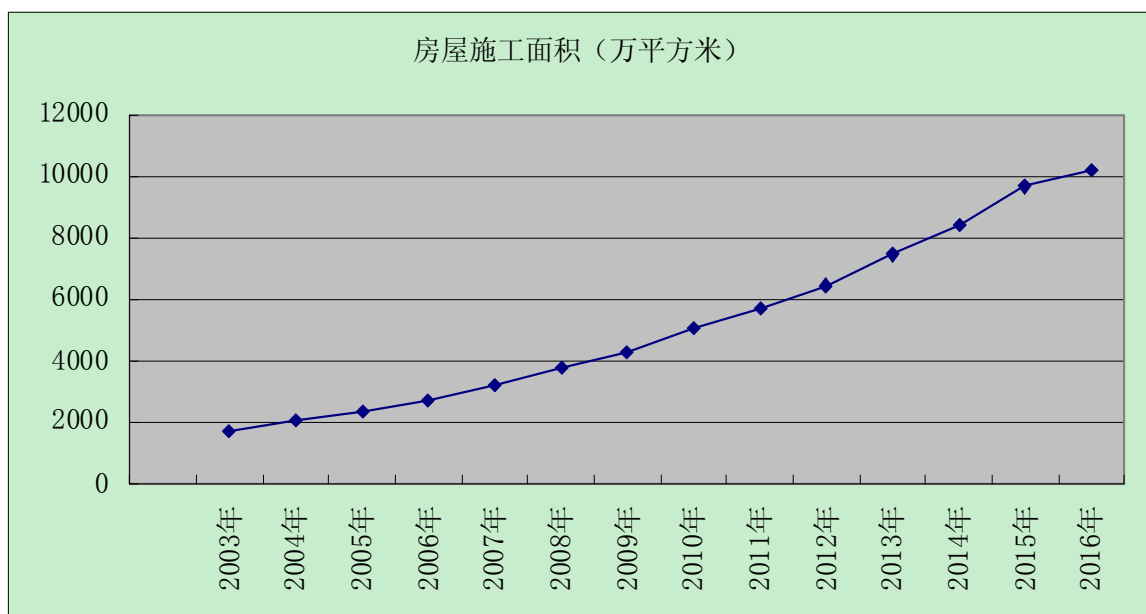


图 5-1 2013-2016 年青岛市房屋施工面积

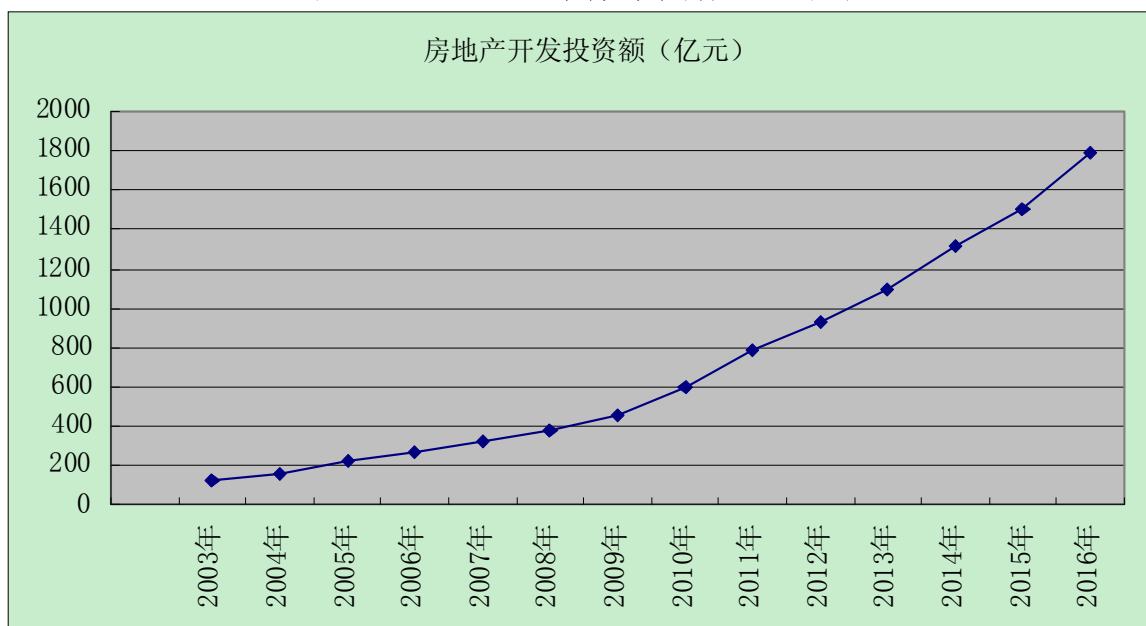


图 5-2 2013-2016 年青岛市房地产开发投资额



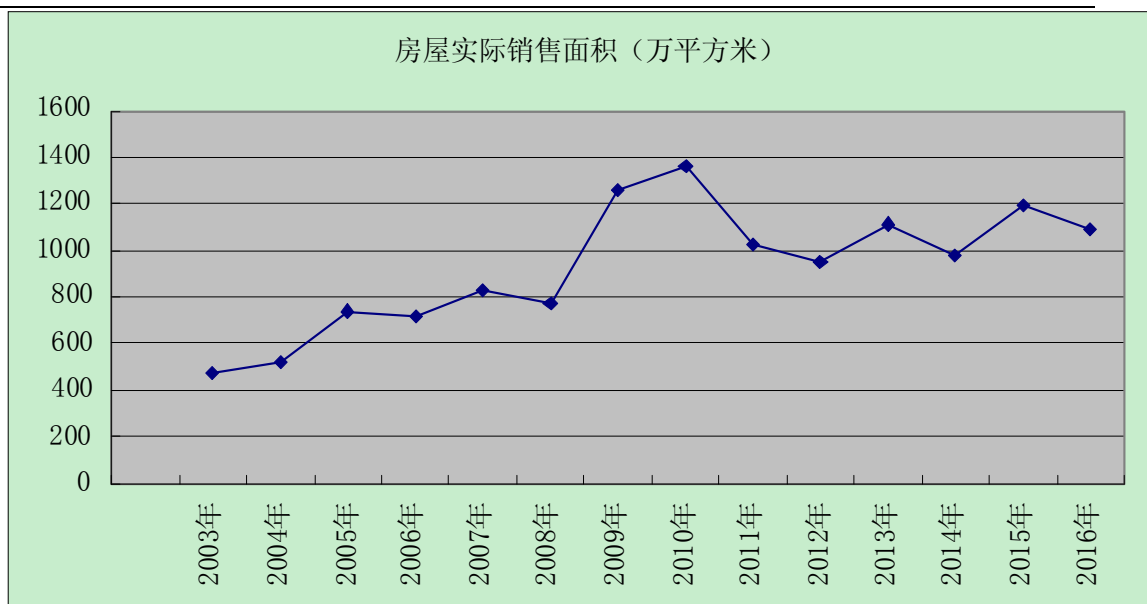


图 5-3 2013-2016 年青岛市房屋实际销售面积

## 5.4 本章小结

本章在详细介绍了组合预测及其意义的基础上，给出了适用于青岛房地产市场组合预测的一种非最优组合预测方法——熵值法。通过运用熵值法把前两章使用的单项预测模型组合起来，通过三种模型的比较发现该组合模型既有趋势预测的优点，又有因素预测的优势。运用该模型对青岛市房地产市场进行预测取得了较好的精度和应用价值。

## 6 结论与展望

### 6.1 结论

本文主要提出了一种对于青岛房地产市场预测的一种新的方法,该方法主要集合了灰色预测和神经网络预测的思想精髓,通过熵值法对上述两种模型进行组合预测。本文所做的工作及得到的结论主要有以下几个方面:

(1) 经过预测青岛市 2013-2016 年的房地产市场相关指标可以看出,青岛市房地产市场能够继续保持稳定繁荣的方向发展,房地产开发商开发规模和投资金额将会逐渐加大,但房屋销售面积的增长趋势可能有所减缓。总结分析了近些年青岛市房地产市场上的一些变化,发现青岛房地产市场总体来说是健康有序的。

(2) 运用改进  $GM(1,1)$ ——马尔柯夫模型对青岛是房地产市场进行组合预测时发现,对于房屋施工面积这一指标取得了极高的预测精度,但对于房地产开发投资额这一指标预测精度就不尽如人意,另外对于房屋实际销售面积该模型直接就无法通过模型检验,这样我们可以发现,对于趋势性很强的指标数据可能运用灰色理论预测比较合适,而对于那些影响因素很复杂的指标往往不能得到预期的效果。

(3) 神经网络方法较灰色预测方法就模型精度这一方面来看的话具有很大的优势。另外通过比较 BP 神经网络和小波神经网络的预测结果我们可以看出,因素分析预测在一般情况下要好于趋势预测方法。总的说来,像神经网络这种新型的智能算法肯定是优于传统的预测方法的。

(4) 组合预测模型能够对预测系统作出比较全面的信息反映,在实验精度上来说并不见得比某种单项预测好,但预测原理较单项预测更为合理。另外如果某个单项预测的实验精度很差,就应舍弃这种单项预测,以免造成不必要的系统误差。

### 6.2 展望

在论文的写作过程中,由于时间仓促以及本人的知识水平有限,论文仍有很多不足和缺陷。我们可以对于以下几个方面作进一步研究和完善:

(1) 文章运用了灰色理论和神经网络理论的组合模型对青岛市房地产市场进行了预测,其实在常用的科学研究中还有很多种预测方法可以运用,本文运用这两种方法取得了较为不错的效果,但是对于青岛市房地产市场其他相关重要指标是否有更好的预测方法这有待以后的同学进行探究。

(2) 文章运用小波神经网络时,对样本数据没有经过算法式的筛选,虽然神经网络有很好的自适应性,但如果通过某种方式选出几个影响因子加大的因素,是不是能简化网络结构,缩短网络训练的时间。

(3) 由于本人对组合预测理论的掌握情况选择了自己熟悉的熵值法,对于该问题是否有更为合适的组合方法有待以后同学进行研究。

(4) 本文选取的模型是从趋势分析和因素分析两个方面进行研究的,对于日常生活中的各种预测性问题我们是否能够从多角度进行分析,而不是仅限于运用相关不同的预测模型,这是值得我们思考的。

相信经过预测理论和方法的不断发展,现实生活中很多迫需解决的问题都会找到其适用的预测方法,从而使各种预测方法在不同的领域中发挥着更为重要的作用。

## 参考文献

- [1]王成庆、况伟大. 1998—2003：中国房地产市场的形式与运作[C].中国房地产发展报告 No.1,北京：社会科学文献出版社,2004.6.89-97
- [2]孟晓苏.中国房地产发展[M].北京:北京大学音像出版社，2007.
- [3]李涛, 伍建平. 房地产泡沫的成因及防范措施探讨[J]. 山东理工大学学报(社会科学版), 2004, 20(2): 32-34
- [4]刘洪玉, 孙峤.房地产价格与房地产经济学研究[J].建筑经济, 2006, (2):12-15
- [5]伊利 (R.T.Ely),莫尔豪斯 (E.W.Morehouse).土地经济学原理[M].中译本商务引书馆,1982
- [6]E.O.Olsen. A Competitive Theory of the Housing Market[J]. American Economic Review, 1969, 159(4):612-622
- [7]Clemen R .T.Linear Constrain and the Efficiency of combined Forecasts[J].Journal of Forecasting,1986.
- [8]Mankiw N G, David N W. The Baby Boom[J].the Baby burst and the Housing Market Regional Science and Urban Economics, 1989(19): 235-258
- [9]Box,G.E.P.,and G.C.Tiao, Comparison of forecast and actuality[J], Appl.Statist,1992, 25: 195-200
- [10]Anthony Y Gu. The predictability of house prices[J].Journal of Real Estate Research, 1994, 24(3):213-234
- [11]Birch John W, Sunderman Mark A. Estimating price paths for residential realestate[J].Journal of Real Estate Research, 1995, 25(3):277-300
- [12]Granger C W J.Ramanathan R.,Improved Methods of Combining Forecasts[J].Journal of Forecasting,1996.
- [13]Palmer PJ. Flouing Development Conferencing London[M]. Chicago:Longman Financial Services Publishing, 1991:83-91
- [14]葛金辉, 桑玉杰. 回归模型在房地产市场方面的应用[J]. 沈阳航空工业学院学报, 2003, 20(4): 69-70
- [15]孙斌艺.现代房地产市场研究理论与方法 [M].上海:上海人民出版社, 2003
- [16]蔡育天. 房地产市场[M]. 上海:同济大学出版社, 1998:45-52
- [17]刘国壁.神经网络新型组合模型在全国能源需求预测中的应用[D].安徽大学:基础数学, 2009.
- [18]Hagan, M.T.等著, 代葵等译.神经网络设计[M].北京:机械工业出版社, 2002:1 — 3.
- [19]陈森林.前向神经网络的分类能力与训练算法的研究[D].西安电子科技大学:计算机应用技术, 2003.
- [20]谭冠军.GM(1,1)模型的背景值构造方法及应用[J].系统工程理论与实践, 2000.5
- [21]赵瑞, 申金山. 住宅需求量预测的灰色马尔柯夫模型及应用[J]. 郑州航空工业管理学院学报 (管理科学版), 2004,(3):21-22
- [22]袁嘉祖. 灰色系统理论及其应用方法[M]. 北京:科学出版社, 1991:43-52
- [23]宋喜民, 周书敬. 房地产市场有效需求分析的灰色数学方法[J]. 河北建筑科技学院学报, 2003, 20(1): 65-68
- [24]曹光辉, 赵骤撰. 重庆市主城区住宅需求总量预测模型[J]. 重庆大学学报,2005, 25(11): 22-25
- [25]唐小我, 经济预测与决策新方法及其应用研究[M], 成都:电子科技大学出版社, 1997 年

- [26]王明涛, 确定组合预测权系数最优近似解的方法研究[J], 系统工程理论与实践, 2000, 20(3):104 — 109
- [27]王秋萍. 组合预测模型的构建及应用[D].西安: 西安理工大学,2008.
- [28]陈俊. 上海市房地产预测系统研究分析[D]. 上海: 同济大学, 2005
- [29]周伟林, 严冀. 城市经济学[M]. 第二版. 上海: 复旦大学出版社, 2004:191-192
- [30]牛凤瑞主编.房地产蓝皮书:2004 年中国房地产发展报告「M].北京:社会科学文献出版社, 2004
- [31]林增杰等. 房地产经济学[M]. 第二版. 北京: 中国建筑工业出版社, 2003:47-49
- [32]傅十和.土地投机与地价泡沫[J].中国土地科学, 1999.2
- [33]梁荣.中国房地产业发展规模与国民经济总量关系研究「M].北京:经济科学出版社, 2005.
- [34]陈则明. 房地产泡沫、房地产热与景气周期的学术观点辨析[J]. 理论探索与争鸣, 2004, 1(2): 22-24
- [35]丹尼斯.迪帕斯奎尔、威廉.C.惠顿, 龙奋杰等译. 城市经济学与房地产市场[M].北京: 经济科学出版社,2002.7
- [36]中国估价师协会. 房地产基本制度和政策[M].北京: 中国物价出版社,2002
- [37]邓聚龙. 灰色系统基本方法[M]. 武汉:华中科技大学出版社, 2005:23-35
- [38]Jenkins, G.M. Relationships between Bsyesian and confidence limits forpredictors[M]. Statist. Soc. , 1964, B26: 176-210
- [39]Raudall Johonston Pozdena. The Housing Principles[M]. Chicago:LongmanFinancial Services Publishing,1987:176-188
- [40]陈洁, 许长新.灰色预测模型的改进[J].辽宁师范大学学报(自然科学版), 2005, 28(3):262 — 264.
- [41]刘思峰, 邓聚龙.GM(1,1)模型的适用范围[J].系统工程理论与实践, 2000, 20(5):121 — 124
- [42]张大海, 江世芳, 史开泉.灰色预测公式的理论缺陷及改进[J].系统工程理论与实践, 2002, 22(8):140 — 142
- [43]王钟羨, 吴春笃.GM(1,1)改进模型及其应用[J].数学的实践与认识, 2003, 33(9):20 — 2
- [44]王沫然.MATLAB 与科学计算[M].北京:电子工业出版社(第二版), 2005.
- [45]Wa.el,H.Awad.Estimating traffic capacity for weaving segments using neural networks technique[J].Applied Soft Computing,2004,(4):395–404.
- [46]Yuan Chang Xie,Yun Long Zhang.A wavelet network model for short-term traffic volume forecasting[J].Journal of Intelligent Transportation Systems,2006,10(3):141-150.
- [47]王莎.BP 神经网络在股票预测中的应用研究[D].中南大学:计算机应用技术, 2008.
- [48]李万庆, 张金水, 孟文清.基于小波神经网络的房地产价格指数预测研究[J].河北工程大学学报(自然科学版), 2008, 25(1):74 — 76
- [49]葛哲学, 孙志强著.神经网络理论与 Matlab r2007 实现[M].北京:电子工业出社, 2007:108 — 110.
- [50]Bates J M,Granger C W J.The Combination of Forecasts[J].Operational Research Quarterly,1969.
- [51]Clemen R T.Linear Constrain and the Efficiency of combined Forecasts[J].Journal of Forecasting,1986.
- [52]马永开, 唐小我, 杨桂元, 非负权重最优组合预测方法的基本理论研究[J], 运筹与管理, 2997, 6(2):z — 8

- [53]张渝, 周宗放, 商业银行信用风险评价指标的熵权选择方法, 电子科技大学学报[J], 2006(6)
- [54]毛开翼.关于组合预测中的权重确定及应用[D].成都理工大学,2007,5.

## 附 录

1997-2010 年青岛人口、经济社会发展主要参数表（小波神经网络预测）

年份		1997	1998	1999	2000	2001	2002	2003
人口	人口密度	2062	2084	2105	2129	2156	2194	2240
	已婚有生育能力的人数	1482204	1505499	1523471	1521331	1514274	1541855	1518448
	社会从业人数（万人）	388.7	393.1	396.1	397.6	400.5	413.31	438.96
	普通高等学校在校人数	27434	29507	33681	46131	60728	82539	168439
经济	人均 GDP	11235	12443	13884	16009	18128	20655	23986
	平均工资	7030	7518	8405	10072	11426	12839	15335
	进出口总额（万美元）	916269	882042	1009885	1353222	1541564	1692567	2065912
	地方财政收入（万元）	476804	580434	680089	800120	987080	1006616	1201398
	金融系统人民币贷款总额	6508776	7239231	8945730	9564293	10798871	13044792	16783917
	居民消费和零售价格指数（1950 年价格为 100）	687.8	687.8	689.2	711.9	719	711.1	721.1
	消费品零售总额（万元）	3008981	3368736	3762754	4282871	4911735	5574364	6455113
社会发展水平	规模以上固定资产投资总额	1611902	1909711	2205891	2426820	2934728	3683623	5475526
	建筑安装工程投资总额（万元）	875057	1034500	1456473	1476397	1781235	2328482	3544790
	年末道路面积（万平方米）	1436	1436	1698	1789	2169	2594	2816
	全年用电量（亿千瓦时）	77.82	79.65	88.12	106.8	116.1	131.444	147.68
	全年客运量（万人次）	55642	51042	55414	59879	60167	61713	58792

年份		2004	2005	2006	2007	2008	2009	2010
人口	人口密度	2229	2290	2338	2377	2384	1873	1873
	已婚有生育能力的人数	1540706	1590476	1604665	1595390	1587344	1623969	1591103
	社会从业人数（万人）	458.81	471.03	490.1	505.8	513.8	525.71	540.34
	普通高等学校在校人数	201739	239761	260339	264917	269314	275157	284788
经济	人均 GDP	28540	33085	38608	44964	52266	57251	65827
	平均工资	17190	20022	23457	27083	30233	33258	37805
	进出口总额（万美元）	2698781	3302230	3911543	4572534	5363659	4485115	5705963
	地方财政收入（万元）	1305136	1763412	2257663	2925798	3424359	3769896	4526138
	金融系统人民币贷款总额	18472613	20394792	25779442	30970553	37483209	48735326	58862263
	居民消费和零售价格指数（1950 年价格为 100）	736.2	753.1	759.9	794.1	831.4	835.6	854
	消费品零售总额（万元）	7475022	8701057	10163457	12162246	14922153	17302231	19611331
社会发展水平	规模以上固定资产投资总额	9845646	14032960	14856894	16353636	20190098	24588889	30224785
	建筑安装工程投资总额（万元）	6413046	8949231	8818614	9981666	11795435	14911874	18725922
	年末道路面积（万平方米）	3254	3548	5218	5411	5596	5763	5893
	全年用电量（亿千瓦时）	167.95	193.81	215.35	236.42	250.07	259.42	292.97
	全年客运量（万人次）	66463	68776	73726	78702	82460	81768	85251



Matlab 程序:

```
%-----改进型灰色预测方法-----

c=[a b]';
L=[ ];
A=L.^(1/2)
B=cumsum(A); % 原始数据累加
n=length(A);
for i=1:(n-1)
    C(i)=(B(i+1)-B(i))/(log(B(i+1))-log(B(i)));
end
% 计算待定参数的值
D=A;D(1)=[ ];
D=D';
E=[-C;ones(1,n-1);[2:n]];
c=inv(E'*E)*E'*D;
c=c';
a=c(1);b1=c(2);b2=c(3);
d=0;f=0;g=0;h=0;
for i=2:n
    d=d+exp(-2*a*(i-1));
    f=f+exp(-1*a*(i-1));
    g=g+i*exp(-1*a*(i-1));
    h=h+B(i)*exp(-1*a*(i-1));
end
v=((b1+b2)/a-b2/(a*a))*d+(b2/(a*a)-b1/a)*f-b2/a*g+h)/(d*A(1))
% 预测后续数据
F=[];F(1)=A(1);
for i=2:(n+6)
    F(i)=(v*A(1)-b1/a-b2/a+b2/(a*a))*exp(-a*(i-1))+b1/a+b2*i/a-b2/(a*a);
end
G=[];G(1)=A(1);
for i=2:(n+6)
    G(i)=F(i)-F(i-1); %得到预测出来的数据
end
%画图
t1=1997:2010;
t2=1997:2016;
T=G;
G=G.^2;
plot(t1,L,'o',t2,G) %原始数据与预测数据的比较
xlabel('年份')
ylabel('房屋施工面积')
% x=[1:5]
% ones(1,6)
```

```
%-----BP 神经网络主要程序-----

input_train=[];
output_train=[];
[inputn,inputps]=mapminmax(input_train);
[outputn,outputps]=mapminmax(output_train);
NodeNum = 59;          % 隐层节点数
TypeEnum = 1;          % 输出维数
Epochs = 10000;        % 训练次数
TF1 = 'tansig';TF2 = 'purelin'; % 设置传递函数
net = newff(minmax(inputn),[NodeNum TypeNum],{TF1 TF2},'trainlm');
net.trainParam.epochs = Epochs;      % 最大训练次数
net.trainParam.goal = 1e-3;          % 最小均方误差
net.trainParam.min_grad = 1e-5;      % 最小梯度
net.trainParam.show = 200;           % 训练显示间隔
net.trainParam.time = inf;           % 最大训练时间
net = train(net,inputn,outputn);     % 训练
input_text = [];
inputn_test=mapminmax('apply',input_text,inputps);
an = sim(net,inputn_test);            % 测试，输出为预测值
BPoutput=mapminmax('reverse',an,outputps);

%-----小波神经网络主要程序-----

%-----
% 产生训练样本与测试样本
input_train=[];
output_train=[];
[inputn,inputps]=mapminmax(input_train);
[outputn,outputps]=mapminmax(output_train);
% 函数接口赋值
NodeNum = 36;          % 隐层节点数
TypeEnum = 1;          % 输出维数
Epochs = 1000;        % 训练次数
input_text = [];
%-----
% 设置网络参数
TF1 = 'template_transfer';TF2 = 'purelin';
```

```
% TF1 = 'radbas';TF2 = 'purelin';
net = newff(minmax(inputn),[NodeNum TypeNum],{TF1 TF2},'trainlm');
% 指定训练参数
net.trainParam.epochs = Epochs;      % 最大训练次数
net.trainParam.goal = 1e-5;           % 最小均方误差
net.trainParam.min_grad = 1e-20;      % 最小梯度
net.trainParam.show = 200;            % 训练显示间隔
net.trainParam.time = inf;            % 最大训练时间
%-----
% 训练与测试
inputn_test=mapminmax('apply',input_text,inputps);
net = train(net,inputn,outputn);      % 训练
X = sim(net,inputn_test);             % 测试 - 输出为预测值
%-----
xiaobooutput=mapminmax('reverse',X,outputps);
```

## 致 谢

本研究是在导师刘强教授的悉心指导下完成的。在三年的学习中，刘老师教给我的不仅仅是知识，更重要的是做人做事的道理，他严谨的治学态度，满腔的工作热情，无畏的探索精神是我终身受益的财富。在此，向敬爱的刘老师表示最诚挚的敬意！

衷心感谢冯启民教授、高惠瑛教授、崔俊山教授、焦双健副教授、赵林副教授、孙燕副教授、潘峰老师对我学习上的教导和生活上的无私帮助；感谢师兄许尧，师姐彭晓晓在学习和生活中对我的建议与帮助；感谢实验室同届同学柏延震、夏书娟、张玉红在论文撰写过程中给予我的帮助与鼓励，感谢实验室田亚力、施建平等师弟师妹们，对于本研究的完成都给予了许多建议和帮助，在此一并表示诚挚的谢意！

感谢多年来一直在家中操劳的父母，他们所给予我的生活上的关心、学习上的鼓励，永远是我生命中的瑰宝，让我无论遭遇何种逆境，都有足够的信心应对！在此，我要向他们表示最深切的尊敬和谢意！

最后，我要感谢中国海洋大学，感谢中国海洋大学土木系，所提供的积极而朴实的学习氛围以及开阔而自由的交流平台，让我可以在这方平台上，接触更多的知识，经过三年的学习和充实，我的视野更开阔，思想更成熟，脚步也更坚实！

## 个人简历

1987 年 9 月 16 日出生于江苏省常州市。

2006 年 9 月考入中国海洋大学数学科学学院信息与计算科学专业，  
2010 年 6 月本科毕业并获得理学学士学位。

2010 年 9 月考入中国海洋大学工程学院管理科学与工程专业，攻读硕士学位至今。

## 发表的学术论文

- [1] 刘强;滕渊;.重大自然灾害灾后重建可持续发展评价模型研究[J]防灾减灾学报.2012,(03)
- [1] 刘强;滕渊;.构建与完善我国巨灾保险体系研究[J]工程地质学报.2012,(20)增刊