

# 基于 BP 神经网络组合模型的中国网络零售额预测研究

刘晓敏<sup>1</sup>, 冯志成<sup>2</sup>, 施建华<sup>3\*</sup>

1. 闽南师范大学数学与统计学院, 福建 漳州 363000

2. 数据科学与统计重点实验室, 福建 漳州 363000

3. 福建省粒计算及其应用重点实验室, 福建 漳州 363000

**摘要:** 网络零售交易规模是反映国内经济繁荣程度、网络零售市场变化以及居民网络消费需求的重要经济指标。为了更好地反映我国网上零售交易规模的总量以及未来的发展趋势, 科学有效地预测网络零售额的未来趋势发展相当重要。本文针对新冠疫情前的2005~2020年中国网络零售额进行预测研究, 考虑 ARIMA、Holt 两参数指数平滑、GM (1,1) 以及主成分回归四种模型, 并基于 BP 神经网络的非线性组合预测方法结合四种单一预测方法得出的预测结果, 最终考虑基于 ARIMA-Holt-GM 的 BP 神经网络组合预测模型, 从而科学预测 2021 年的中国网络零售额, 为政府和电商企业的科学合理决策提供一种更可靠的统计分析方法。

**关键词:** 网络零售额; BP 神经网络; 组合模型

## Research on Predicting China's Online Retail Sales Based on BP Neural Network of Combination Model

Liu Xiaomin<sup>1</sup>, Feng Zhicheng<sup>2</sup>, Shi Jianhua<sup>3\*</sup>

1. School of Mathematics and Statistics, Minnan Normal University, Zhangzhou, Fujian 363000

2. Fujian Key Laboratory of Data Science and Statistics, Zhangzhou, Fujian 363000

3. Fujian Key Laboratory of Granular Computing and Applications, Zhangzhou, Fujian 363000

**Abstract:** The scale of online retail transactions is an important economic indicator that reflects the level of domestic economic prosperity, changes in the online retail market, and residents' demand for online consumption. In order to better reflect the total volume of online retail transactions and future development trends in China, it is crucial to scientifically and effectively predict the future trend of online retail sales. This paper conducts prediction research on China's online retail sales from 2005 to 2020 before the COVID-19, considering four models of ARIMA, Holt two parameter exponential smoothing, GM (1,1) and principal component regression, and combining the prediction results obtained by the nonlinear combination prediction method based on BP neural network with four single prediction methods, finally considering the combination prediction model based on ARIMA Holt GM, so as to scientifically predict China's online retail sales in 2021, and provide a more reliable statistical analysis method for the government and e-commerce enterprises to make scientific and reasonable decisions.

**Keywords:** online retail sales; BP neural network; combination model

## 引言

随着中国互联网普及率的上升以及信息技术的快速发展, 越来越多的网上购物平台出现在市场上, 由于网络购物的发展大量节约了市场沟通的成本, 买卖双方的交易也不再受时间和空间上的限制, 以及琳琅满目的商品种类和方便快捷又灵活的物流服务等特点, 在很大程度上方便了人们的生活, 因此人们越来越依赖于互联网络来从事商品的买卖活动, 消费习惯也从“线下”慢慢地转移到“线上”, 网络购物行为也由此得到快速发展和进步, 网络购物已经成为我国居民消费的重要渠道, 给中国消费经济注入了新鲜血液。依托 5G、云计算、大数据等为代表的新一代信息技术的飞速发展, 以及一系列利好政策的持续发力: 推动实体经济转型、发展跨境电商、支持农村电商<sup>[1]</sup>, 让网络购物市场迎来了爆发式增长, 其交易规模呈现大幅增长态势, 在社会消费品零售总额中的占比也是连年提升, 为我国经济实现高质量发展发挥着重要作用。

基金项目: 国家社会科学基金 (20XTJ003)。

作者简介:

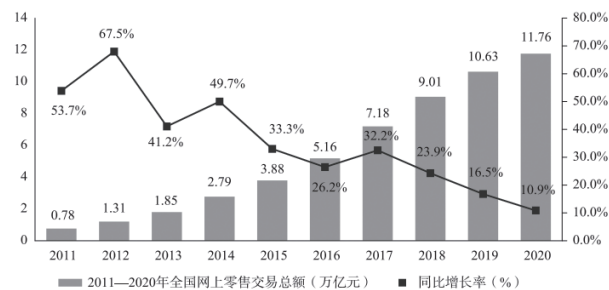
刘晓敏 (1999—), 女, 福建福州人, 学士, 专业方向为应用统计;

冯志成 (2001—), 男, 福建平潭, 硕士研究生, 专业方向为概率统计。

通讯作者:

施建华 (1977—), 男, 福建福清人, 闽南师范大学数学与统计学院, 教授, 博士, 硕士生导师, 研究方向: 统计理论与方法, 邮箱: v0085@126.com。

据中华人民共和国商务部电子商务和信息化司发布的《2020年中国电子商务发展总报告》显示,新冠肺炎疫情发生以来,网络零售市场仍然保持着平稳增长的势头,2020年全国网上零售额达11.76万亿元,同比增长10.9%(见图1-1),其中实物商品的网上零售额达到9.76万亿元,同比增长14.8%,在社会消费品零售总额中占24.9%,比上年提高4.2个百分点,成为消费市场的稳定器<sup>[2]</sup>。“十三五”期间,即2016~2020年,全国网上零售额从5.16万亿元增长到11.76万亿元,年均增长率为22.9%,网络零售市场规模的逐年增长趋势使得电商主体在我国经济发展中的地位得到了一定的提升。



资料来源:中华人民共和国商务部,《中国电子商务报告(2020)》,2021年9月15日  
>图1-1 2011-2020年全国网上零售交易规模

2021年是“十四五”规划的开局之年,也是我国电子商务处在由高速发展向高质量发展全面转型的关键时期<sup>[2]</sup>,而网络零售交易规模(网络零售额)是反映国内经济繁荣程度、网络零售市场变化以及居民网络消费需求的重要经济指标<sup>[1]</sup>。因此,有效科学地预测网络零售额的未来发展趋势,有助于政府制定相关政策,以促进居民消费、提升国家经济发展水平,有利于电商企业明确网络零售市场容量,准确定位自身市场地位,优化资源配置,不断调整产业结构,提高产品品质、提升服务质量,也有利于揭示网络零售市场潜在需求与机会,为战略投资者选择合适的投资时机提供准确的市场情报信息,为公司领导层进行战略规划提供科学的决策依据,从而进一步促进网络零售交易规模的增长,促进电子商务由高速发展向高质量发展全面转型。

综上所述,网络零售额的预测对政府、电商企业和市场相当重要。如何选择科学合理的预测方法使得预测结果更加准确,并根据预测结果为市场提供科学的决策依据,这是本文的主要研究课题。本文将首先采用ARIMA、Holt两参数指数平滑、GM(1,1)以及主成分回归(PCR)四种模型对新冠疫情前的2005~2020年中国网络零售额进行分析并单一预测(由于篇幅所限没有详细展示建模预测过程),再利用基于BP神经网络的组合预测方法结合四种单一预测方法得出的预测结果,综合之后,得到基于ARIMA-Holt-GM的组合预测模型,最后对比分析不同模型预测效果的精确度,得出较为准确的2021年中国网络零售额预测值。

## 一、网络零售额预测的相关模型简介

### (一) ARIMA模型

众所周知,时间序列预测法是一种定量预测方法,其中,较常用的两种时间序列模型分别是自回归移动平均模型(ARMA)和自回归差分移动平均模型(ARIMA),前者的适用范围是原始序列数据具有平稳性且不存在差分阶数,则可以考虑使用ARMA模型进行建模预测;后者是当原始序列数据具有不平稳性且需要进行差分处理时,则可以考虑使用ARIMA模型<sup>[3][4]</sup>。在本文研究中,由于选取的2005~2020年中国网络零售额的年度数据并不具有季节性,而且网络零售额呈现逐年上升的趋势,明显为非平稳序列,故可以采用自回归差分移动平均模型(ARIMA)来对中国网络零售额来进行预测。

### (二) Holt两参数指数平滑模型

若在采用时间序列中的自回归差分移动平均模型(ARIMA)时,发现序列需要多次差分后才能达到平稳,那么此时我们可以考虑采用Holt两参数指数平滑法<sup>[5]</sup>来进行预测,从而解决因过差分而损失大量信息并导致预测精度降低的问题。在本文研究中,考虑到中国网络零售额的历史数据并不具有季节性,且网络零售额呈现逐年上升的线性趋势,符合Holt两参数指数平滑法对建模数据的条件,故可以采用指数平滑法中的Holt两参数指数平滑法对中国网络零售额进行预测。

### (三) GM(1,1)模型

上述模型对数据量有一定的要求,当数据量太少时,其建

立的预测模型也有可能因为样本数据量的不足从而影响到预测精度。于是,本文在此基础上考虑添加采用灰色预测法中的GM(1,1)灰色预测模型,即1阶方程、1个变量的微分方程型的灰色模型,这是目前应用最广泛的灰色预测模型。它适用于解决信息量少,数据近似呈指数规律变化的时间序列。在本文中,影响中国网络零售额的因素很多,部分因素已知,部分因素未知,因此可以将网络零售额视为灰色系统,建立灰色预测模型GM(1,1),对中国网络零售额进行预测。

### (四) BP神经网络模型

神经网络是模拟生物神经网络进行信息处理的一种数学模型,是一个具有学习能力的系统。它可以建立某种映射关系(函数),并进行函数的拟合,即通过看似杂乱无章的大量历史数据中发现规律,以此获得有效的预测。在分类预测时,神经网络普遍采用有指导的学习方式,通过给定的训练样本来调整神经网络的参数,使得网络的输出结果接近于已知的样本类标记。目前,最广泛应用的人工神经网络是采用前馈反向传播网络,即BP神经网络,它是前向网络的核心,体现了神经网络最精华的部分。本文将采用BP神经网络方法对网络零售额进行组合预测,并与其他单一预测方法进行比较。

### (五) 组合模型

组合预测方法是针对同一个问题,采用两种或两种以上不同单一预测方法的组合预测。它既可以是几种定量方法的组合,也可是几种定性的方法的组合,或者定性和定量的组合,组合预测

的主要目的是综合利用各种单一预测方法所提供的信息，尽可能地提高预测的精度。各单一预测模型在对数据进行预测时，可能会由于自身条件的限制，不能全面掌握数据的信息，从而影响预测结果<sup>[6]</sup>，预测结果的精确度也因此而降低，且预测的风险也相对较大。因此，可以采用组合预测的方法，通过一定规则组合各单一预测模型，对各单一预测模型进行加权平均，利用单一预测模型间的优势互补，从而达到优化预测结果的作用。

组合预测模型的重点在于加权系数的确定。如果对各个单一预测模型的加权系数赋值合理，那么整个组合预测模型的预测精度也会相应提高<sup>[9]</sup>，相关学者的研究均证实，基于神经网络的非线性组合预测方法具有优良的特性和更高的预测精度<sup>[7]</sup>。因此，本文将采用BP神经网络方法来对各单一预测模型进行组合预测，即通过ARIMA模型、Holt两参数指数平滑模型、GM(1,1)模型和PCR模型的预测结果进行加权平均，得到最终的预测结果。

借鉴彭乃驰，党婷<sup>[8]</sup>的思路，本文考虑的ARIMA-Holt-GM-PCR组合预测模型的具体步骤为：

(1) 使用ARIMA模型作单一预测得到预测值，其中， $\hat{y}_i$ 表示为第*i*年的预测值；

(2) 使用Holt两参数指数平滑模型作单一预测，得到的预测值为 $\hat{y}_i$ ，其中， $\hat{y}_i$ 表示为第*i*年的预测值；

(3) 使用GM(1,1)模型作单一预测得到预测值，其中， $\hat{y}_i$ 表示为第*i*年的预测值；

(4) 使用PCR模型作单一预测得到预测值，其中， $\hat{y}_i$ 表示为第*i*年的预测值；

(5) 确定BP神经网络的隐层和输出层传递函数、训练函数、学习函数、表现函数及训练参数；

(6) 以 $x_1, \dots, x_m$ 为BP神经网络输入训练样本( $m < n$ )，以为输出目标，其中表示为第*i*年的真实值；

(7) 根据经验公式( $a$ 为输入节点数， $b$ 为输出节点数， $c$ 为1-10的整数)初步确定隐层神经元个数*s*的取值范围；

(8) 通过训练数据创建BP神经网络模型并保存该网络；

(9) 输入测试样本 $x_1, \dots, x_m$ ，通过保存的网络预测出测试年份的预测值，并与其真实值进行对比，检验测试网络的预测效果；

(10) 利用建立的网络对未来年份进行预测。

## 二、网络零售额预测指标构建

### (一) 网络零售额影响因素分析

本文在查找与相关网络零售额的影响因素时，结合数据的可获得性、可操作性、系统性和客观性原则，将网络零售额的影响因素主要划分为宏观经济因素、市场规模因素和用户规模因素三个方面，一共整理了15个与网络零售额密切相关的因素指标，具体如下：国内生产总值、城镇居民人均可支配收入、居民消费水平（归类为宏观经济因素）；互联网普及率、快递业务量、互联网数据中心市场规模、B2C市场交易规模、中国移动支付交易规模、B2C和C2C电子商务企业数量、第三方移动支付市场交易规模、移动购物市场交易规模（归类为市场规模因素）；网络购物用户规模、网络

支付用户规模、互联网宽带接入用户规模、女性网购用户规模（归类为用户规模因素）。但由于部分数据的不可获得性，或者数据可获得但存在大量缺失值，若仍将此部分数据纳入指标体系，会在一定程度上影响模型拟合的准确性。因此，本文将从中选择数据比较完整的9个影响因素指标，进行下一步的相关性分析。

表2-1 指标体系及其数据来源

影响因素	变量	指标	单位	数据来源
	W	网络零售额	亿元	中国电子商务报告
宏观经济	X1	国内生产总值	亿元	国家统计局
	X2	城镇居民人均可支配收入	元	中国统计年鉴
	X3	居民消费水平	元	国家统计局
市场规模	Y1	互联网普及率	%	国家统计局
	Y2	快递业务量	万件	国家统计局
	Y3	互联网数据中心(IDC)市场规模	亿元	智研咨询、中国IDC圈及赛迪顾问
	Y4	B2C市场交易规模	亿元	Analysys易观、易观国际、易观智库
用户规模	Z1	网络购物用户规模	万人	中国电子商务报告
	Z2	网络支付用户规模	万人	中国互联网络信息中心、中国网络支付安全状况报告、艾瑞咨询

表2-2 2005-2020年全国网络零售额与宏观经济因素数据

年份	网络零售额 / 亿元	国内生产总值 / 亿元	城镇居民人均可支配收入 / 元	居民消费水平 / 元
2005	150	187318.9	10493	5688
2006	280	219438.5	11760	6319
2007	620	270092.3	13786	7454
2008	1257	319244.6	15781	8505
2009	2586	348517.7	17175	9249
2010	5091	412119.3	19109	10575
2011	7826	487940.2	21810	12668
2012	13110	538580	24565	14074
2013	18636	592963.2	26467	15586
2014	27898	643563.1	28844	17220
2015	38773	688858.2	31195	18857
2016	51556	746395.1	33616	20801
2017	71751	832035.9	36396	22969
2018	90065	919281.1	39251	25245
2019	106324.2	986515.2	42359	27504
2020	117601.3	1015986.2	43834	27438

### (二) 网络零售额指标建立

#### 1. 指标数据获取

本文以2005年至2020年各影响因素数据为研究区间，结合相关渠道给出的数据，查找并整理了宏观经济、市场规模和用户规模三个方面的9个影响网络零售额因素的指标数据。具体指标体系及其数据来源如表2-1所示，具体数据如附录中的表2-2、2-3、2-4所示。

表2-3 2005-2020年全国网络零售额与市场规模因素数据

年份	网络零售额 / 亿元	互联网普及率 %	快递业务量 / 万件	互联网数据中心 (IDC) 市场规模 / 亿元	B2C 市场交易规模 / 亿元
2005	150	8.5	22880.3	13.8	20
2006	280	10.5	26988	21.6	28
2007	620	16	120189.6	34.6	43
2008	1257	22.6	151329.3	48.7	81.2
2009	2586	28.9	185786	72.8	220.2
2010	5091	34.3	233892	102.2	1028.9
2011	7826	38.3	367311	170.8	2235.5
2012	13110	42.1	568548	210.8	4639
2013	18636	45.8	918674.89	262.5	8283
2014	27898	47.9	1395925.3	372.2	13701.1
2015	38773	50.3	2066636.84	518.6	20136.3
2016	51556	53.2	3128315.11	714.5	27392.4
2017	71751	55.8	4005591.91	946.1	38284.3
2018	90065	59.6	5071042.8	1228	50164.4
2019	106324.2	-	6352290.97	1562.5	61327.4
2020	117601.3	70.4	8335789.43	1958.2	72600.7

表2-4 2005-2020年全国网络零售额与用户规模因素数据

年份	网络零售额 / 亿元	网络购物用户规模 / 万人	网络支付用户规模 / 万人
2005	150	1855	600
2006	280	3357	1600
2007	620	4641	3300
2008	1257	7400	5200
2009	2586	10800	9406
2010	5091	16051	13719
2011	7826	19395	16676
2012	13110	24202	22065
2013	18636	30189	26020
2014	27898	36142	30431
2015	38773	41325	41618
2016	51556	46670	47450
2017	71751	53332	53110
2018	90065	61011	60004
2019	106324.2	-	-
2020	117601.3	78241	85434

## 2. 数据预处理及相关性分析

由于数据来源缺失2019年度的互联网普及率、网络购物用户规模及网络支付用户规模相关数据,为保证数据的连续性、完整性以及后续相关性分析的可行性,本文将采用牛顿插值法对缺失数据进行插补处理,插补结果显示,插补后的数据基本符合数据的增长趋势,因此,本文将采取插补后的数据对其进行相关性分析,插补后数据如附录中的表2-5、2-6所示。

表2-5 市场规模因素插补后数据

年份	网络零售额 / 亿元	互联网普及率 %	快递业务量 / 万件	互联网数据中心 (IDC) 市场规模 / 亿元	B2C 市场交易规模 / 亿元
2005	150	8.5	22880.3	13.8	20
2006	280	10.5	26988	21.6	28
2007	620	16	120189.6	34.6	43
2008	1257	22.6	151329.3	48.7	81.2
2009	2586	28.9	185786	72.8	220.2
2010	5091	34.3	233892	102.2	1028.9
2011	7826	38.3	367311	170.8	2235.5
2012	13110	42.1	568548	210.8	4639
2013	18636	45.8	918674.89	262.5	8283
2014	27898	47.9	1395925.3	372.2	13701.1
2015	38773	50.3	2066636.84	518.6	20136.3
2016	51556	53.2	3128315.11	714.5	27392.4
2017	71751	55.8	4005591.91	946.1	38284.3
2018	90065	59.6	5071042.8	1228	50164.4
2019	106324.2	65.4	6352290.97	1562.5	61327.4
2020	117601.3	70.4	8335789.43	1958.2	72600.7

表2-6 用户规模因素插补后数据

年份	网络零售额 / 亿元	网络购物用户规模 / 万人	网络支付用户规模 / 万人
2005	150	1855	600
2006	280	3357	1600
2007	620	4641	3300
2008	1257	7400	5200
2009	2586	10800	9406
2010	5091	16051	13719
2011	7826	19395	16676
2012	13110	24202	22065
2013	18636	30189	26020
2014	27898	36142	30431
2015	38773	41325	41618
2016	51556	46670	47450
2017	71751	53332	53110
2018	90065	61011	60004
2019	106324.2	69168	69295
2020	117601.3	78241	85434

在对中国网络零售额进行预测分析前,需要对选取的影响因素指标与网络零售额之间进行相关性分析,从前述影响因素指标中选取与网络零售额相关程度较高的作为预测指标。本文先使用 SPSS 软件对各影响因素计算相关系数,由分析结果显示,所有的指标与网络零售额都是呈现正相关关系,而且相关性极强,除了互联网普及率外,其他指标的相关系数均大于0.9,即所有指标都与网络零售额强相关,因此,可以纳入所有影响因素指标对网

络零售额进行预测分析。

### 三、实证分析

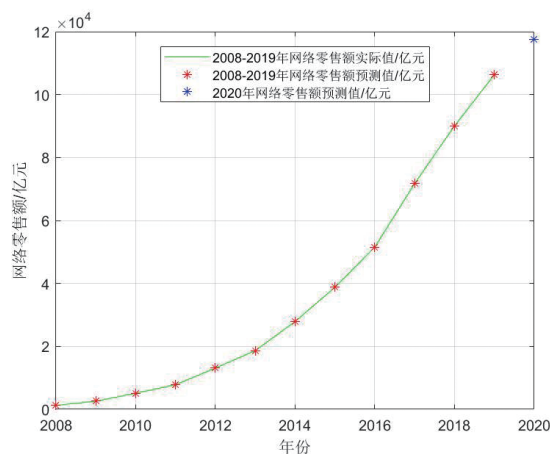
#### (一) 基于BP神经网络的网络零售额组合预测

本文收集了2005~2020年我国网络零售额的数据,但综合考虑了灰色GM(1,1)模型、主成分回归模型(PCR)的拟合效果不佳因素,只选用了2008至2019年的网络零售额作为训练数据。因此,本文将选取ARIMA(2,4,0)、Holt两参数指数平滑、GM(1,1)及PCR四个模型的网络零售额预测数据,

作为基于BP神经网络的组合预测的训练数据,将四个模型的2020年的预测数据用于测试模型,最后用训练出的模型预测2021年的网络零售额。本文BP神经网络的组合预测将采用Matlab软件编写建模程序,并进行相应的分析。

基于ARIMA、Holt、GM(1,1)、PCR四个模型的组合预测操作如下:将2008年至2019年四个模型的预测数据作为训练数据,利用MATLAB软件中premnmx函数对样本进行归一化处理,然后对其进行训练与预测,设定网络的最大学习迭代次数50000次,学习精度为 $1e-6$ ,学习速率为0.01,最终得到基于四个模型的BP神经网络训练数据以及测试数据拟合结果(表3-1、3-2),预测值与实际值对比如图3-1所示。

由训练数据拟合效果对比表(表3-1)可以看出,BP神经网络的预测结果非常准确,2008年至2019年预测值的相对误差均低于3%,平均误差(相对误差取绝对值后平均)仅为0.64%,说明BP神经网络模型拟合效果非常好,达到了较高的要求,可以用来预测未来趋势。由BP神经网络的测试数据拟合效果对比表(表3-2)显示,测试数据2020年的预测值的相对误差仅为-0.13%,说明测试效果也非常好,且由中国网络零售额组合预测拟合对比图(图3-1)也可以看出,预测值几乎与实际值曲线完全重合,模



>图1-1 基于四个模型的中国网络零售额组合预测拟合对比图

型预测2020年网络零售额的结果也基本符合指数递增的趋势,说明BP神经网络模型拟合效果和测试效果都达到了很高的要求,因此可以通过建立的BP神经网络模型来预测2021年的中国网络零售额。

考虑到2021年度的几个影响因素指标可获得性,主成分回归模型暂时无法预测2021年的网络零售额。而由上述BP神经网络建模的结果证明,完全可以通过建立BP神经网络模型来预测未来趋势,且精确度有一定的保证。因此,本文将采用ARIMA、Holt两参数指数平滑和GM(1,1)三个模型的网络零售额预测数据作为训练数据来重新建立BP神经网络模型,将三个模型的2020年和2021年的预测数据用于测试模型,从而得出2021年的网络零售额预测值。

基于ARIMA、Holt、GM(1,1)三个模型的组合预测操作如下:将ARIMA、Holt两参数指数平滑和GM(1,1)三个模型的2008年至2019年预测数据作为训练数据,利用MATLAB软件中premnmx函数对样本进行归一化处理,然后对其进行训练与预测,同样设定网络的最大学习迭代次数50000次,学习精度为 $1e-6$ ,学习速率为0.01,最终得到基于三个模型的BP神经网络训练数据以及测试数据拟合结果(表3-3、3-4),预测值与实际值对比如图3-2所示。

表3-3 基于三个模型的BP神经网络训练数据拟合效果对比表

年份	实际值(亿元)	预测值(亿元)	相对误差(%)	平均误差(%)
2008	1257	1293.31148	2.89%	0.79%
2009	2586	2661.06377	2.90%	
2010	5091	5000.802475	-1.77%	
2011	7826	7931.026012	1.34%	
2012	13110	13141.89505	0.24%	
2013	18636	18640.19267	0.02%	
2014	27898	27938.61429	0.15%	
2015	38773	38807.70539	0.09%	
2016	51556	51577.73976	0.04%	
2017	71751	71764.45464	0.02%	
2018	90065	90078.08228	0.01%	
2019	106324.2	106328.4329	0.00%	

由训练数据拟合效果对比表(表3-3)可以看出,基于三个模型的BP神经网络组合预测结果也非常准确,2008年至2019年

表3-1 基于四个模型的BP神经网络训练数据拟合效果对比表

年份	实际值(亿元)	预测值(亿元)	相对误差(%)	平均误差(%)
2008	1257	1264.87938	0.63%	0.64%
2009	2586	2663.418051	2.99%	
2010	5091	4950.221155	-2.77%	
2011	7826	7884.185851	0.74%	
2012	13110	13162.33198	0.40%	
2013	18636	18611.92823	-0.13%	
2014	27898	27888.47805	-0.03%	
2015	38773	38783.60955	0.03%	
2016	51556	51554.7732	0.00%	
2017	71751	71746.93989	-0.01%	
2018	90065	90063.94824	0.00%	
2019	106324.2	106326.6741	0.00%	

表3-2 基于四个模型的BP神经网络测试数据拟合效果对比表

年份	实际值(亿元)	预测值(亿元)	相对误差(%)
2020	117601.3	117450.7413	-0.13%

预测值的相对误差也均低于3%，平均误差为0.79%，说明基于三个模型的BP神经网络组合预测模型拟合效果也非常好，因此也可以用来预测未来趋势。由BP神经网络的测试数据拟合效果对比表（表3-4）显示，测试数据2020年的预测值的相对误差为1.73%，说明测试效果也非常好，另外由中国网络零售额组合预测拟合对比图（图3-2）也可以看出，模型预测2020年和2021年网络零售额的结果也基本符合指数递增的趋势，因此可以通过建立的基于三个模型的BP神经网络组合模型来预测2021年的中国网络零售额，预测值为127389.8389亿元。

表3-4 基于三个模型的BP神经网络测试数据拟合效果对比表

年份	实际值(亿元)	预测值(亿元)	相对误差(%)
2020	117601.3	119640.1873	1.73%
2021	-	127389.8389	-

### (二) 不同模型的预测结果分析对比

现将ARIMA(2,4,0)、Holt两参数指数平滑、GM(1,1)、PCR四个单一模型的预测结果与基于三个模型的BP神经网络的组合预测结果进行对比分析，如表3-5、3-6、3-7、3-8所示。

由2020年的拟合效果对比表（表3-5、3-6）和测试效果对比表（表3-7、3-8）可以看到，不管是拟合效果还是测试效果，组合模型的预测误差基本上都小于单一预测模型的预测误差，除

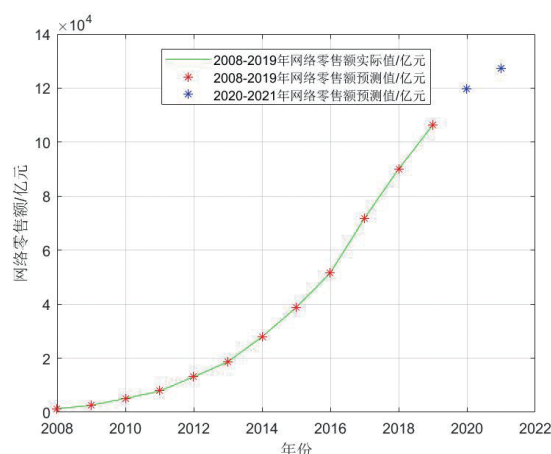


图3-2 基于三个模型的中国网络零售额组合预测拟合对比图

了GM(1,1)模型在2020年的预测误差，组合预测模型在其他年份的预测误差都低于单个模型的预测误差。由此可见，基于ARIMA-Holt-GM组合预测模型的预测结果是相对有效的，组合预测的精度在总体上要优于其他单一预测模型的精度。因此，得出最终2021年的中国网络零售额预测值为127389.8389亿元，而经过查证，2021年度国家统计局给出的全国网上零售额为130884亿元，预测的相对误差为-2.67%。

表3-5 ARIMA和Holt两模型与组合模型拟合效果对比表

年份	实际值(亿元)	ARIMA		Holt		组合模型	
		预测值(亿元)	相对误差(%)	预测值(亿元)	相对误差(%)	预测值(亿元)	相对误差(%)
2008	1257	1257.0936	0.01%	959.68	-23.65%	1293.31148	2.89%
2009	2586	2444.1151	-5.49%	1893.7	-26.77%	2661.06377	2.90%
2010	5091	4914.0978	-3.47%	3914.17	-23.12%	5000.802475	-1.77%
2011	7826	8907.2476	13.82%	7594.72	-2.96%	7931.026012	1.34%
2012	13110	11451.7532	-12.65%	10561.53	-19.44%	13141.89505	0.24%
2013	18636	20566.7322	10.36%	18390.28	-1.32%	18640.19267	0.02%
2014	27898	24932.8660	-10.63%	24163.54	-13.39%	27938.61429	0.15%
2015	38773	41255.6542	6.40%	37154.48	-4.17%	38807.70539	0.09%
2016	51556	51124.3107	-0.84%	49648.34	-3.70%	51577.73976	0.04%
2017	71751	68162.0760	-5.00%	64337.3	-10.33%	71764.45464	0.02%
2018	90065	96833.5798	7.52%	91936.11	2.08%	90078.08228	0.01%
2019	106324.2	110428.9499	3.86%	108387.44	1.94%	106328.4329	0.00%

表3-6 GM(1,1)和PCR两模型与组合模型拟合效果对比表

年份	实际值(亿元)	GM(1,1)		PCR		组合模型	
		预测值(亿元)	相对误差(%)	预测值(亿元)	相对误差(%)	预测值(亿元)	相对误差(%)
2008	1257	1257	0.00%	-1445.880936	-215.03%	1293.31148	2.89%
2009	2586	-7843	-403.28%	3150.099325	21.81%	2661.06377	2.90%
2010	5091	-1226	-124.09%	9241.086891	81.52%	5000.802475	-1.77%
2011	7826	6096	-22.10%	15916.39084	103.38%	7931.026012	1.34%
2012	13110	14200	8.31%	22553.56669	72.03%	13141.89505	0.24%
2013	18636	23168	24.32%	29480.41272	58.19%	18640.19267	0.02%
2014	27898	33092	18.62%	37294.83795	33.68%	27938.61429	0.15%
2015	38773	44075	13.68%	46909.13846	20.98%	38807.70539	0.09%
2016	51556	56230	9.07%	57340.42604	11.22%	51577.73976	0.04%
2017	71751	69681	-2.88%	69291.22966	-3.43%	71764.45464	0.02%
2018	90065	84567	-6.10%	82921.6619	-7.93%	90078.08228	0.01%
2019	106324.2	101041	-4.97%	97908.13468	-7.92%	106328.4329	0.00%

表3-7 ARIMA 和 Holt 两模型与组合模型测试效果对比表

年份	实际值 (亿元)	ARIMA		Holt		组合模型	
		预测值(亿元)	相对误差(%)	预测值(亿元)	相对误差(%)	预测值(亿元)	相对误差(%)
2020	117601.3	121397.61	3.23%	122585.14	4.24%	119640.1873	1.73%
2021	130884	127696.16	-2.44%	138844.53	6.08%	127389.8389	-2.67%

表3-8 GM(1,1)和 PCR 两模型与组合模型测试效果对比表

年份	实际值 (亿元)	GM(1,1)		PCR		组合模型	
		预测值(亿元)	相对误差(%)	预测值(亿元)	相对误差(%)	预测值(亿元)	相对误差(%)
2020	117601.3	119273	1.42%	113021.9285	-3.89%	119640.1873	1.73%
2021	130884	139449	6.54%	-	-	127389.8389	-2.67%

#### 四、结论

本文从不同角度考虑了 ARIMA 模型、Holt 两参数指数平滑模型、GM(1,1)模型、PCR 以及基于 BP 神经网络方法的组合预测模型,分别对中国网络零售额进行了预测分析,其中 Holt 两参数指数平滑模型弥补了 ARIMA 模型因过差分而损失大量信息并导致预测精度降低的问题,GM(1,1)模型弥补了 ARIMA 及 Holt 两参数指数平滑两种模型可能因样本数据量的不足而影响到预测精度的问题,而 PCR 弥补了前三种模型无法反映预测变量与其主要影响因素之间的定量关系的问题,四种单一模型相互取长补短。本文提出的组合模型正是结合了所有单项预测方法包含的有用信息相互取长补短,进而提高预测结果的可靠性和精确性。最后将四种单一预测模型和 ARIMA-Holt-GM 的 BP 神经网络组合模型的预测结果进行对比。本文研究得出以下结论:

(1) 中国网络零售额受到的影响因素众多,是宏观因素和微观因素共同作用的结果,本文首先采用主成分回归分析,选取 9 个影响因素指标来揭示网络零售额与其影响因素之间的关系,反映出网络零售额变化的原因,从而弥补了 ARIMA、Holt 两参数指数平滑和 GM(1,1)三种模型只能依据时间变化的规律来预测变量未

来变化趋势的不足。

(2) 就单一预测模型而言,2008-2020 年间的预测相对误差绝对值平均方面,时间序列 ARIMA 模型要比其他三种模型的预测精度高,ARIMA 模型的平均为 3.23%,其次是 Holt 两参数指数平滑模型的平均为 11.07%,GM(1,1)模型和主成分回归模型的平均超过 50%,而基于 ARIMA-Holt-GM 的 BP 神经网络组合预测模型的平均绝对误差仅为 0.79%。而在分析的过程中发现,如果信息完整的情况下,ARIMA-Holt-GM-PCR 的 BP 神经网络组合预测模型在预测性能上会更加精准。

(3) 不管是拟合效果还是测试效果,基于 ARIMA-Holt-GM 的 BP 神经网络组合预测模型在整体上的预测精度都比 ARIMA 模型、Holt 两参数指数平滑模型、GM(1,1)模型、主成分回归模型要高,组合模型在大部分年份中的预测误差绝对值都要低于单个模型的预测误差,因此根据本文实证研究证明,在新冠疫情前的我国社会经济环境情况下,基于 ARIMA-Holt-GM 的 BP 神经网络组合预测模型的预测结果是相对有效的,组合预测的精度在总体上要优于其他单一预测模型的精度,用组合模型来对中国网络零售额进行预测是科学合理的。

#### 参考文献

- [1] 夏欢. 我国网上零售影响因素的实证研究——基于灰色关联分析模型[J]. 调研世界, 2020(09): 36-41.
- [2] 中华人民共和国商务部. 中国电子商务报告(2020)[DB/CD].20211022182630164.pdf(mofcom.gov.cn).
- [3] 刘新民,王琪,孙秋霞. 基于 Fisher 聚类的公交客流量时间序列预测及对比[J]. 山东科技大学学报(自然科学版), 2019(02): 74.
- [4] 李志超,刘升. 基于 ARIMA 模型、灰色模型和回归模型的预测比较[J]. 统计与决策, 2019(23): 38.
- [5] 林挺葵,等. 粤西地区及各地级市新型冠状病毒肺炎疫情发展趋势预测分析—基于 Holt 双参数指数平滑模型的研究[J]. 实用心脑血管病杂志, 2020, 28(02):13-17.
- [6] 翟静,曹俊. 基于时间序列 ARIMA 与 BP 神经网络的组合预测模型[J]. 统计与决策, 2016(04): 29-32.
- [7] 凌立文,张大斌. 组合预测模型构建方法及其应用研究综述[J]. 统计与决策, 2019(01): 18-21.
- [8] 彭乃驰,党婷. 基于 ARMA-GM-BP 组合预测模型及应用[J]. 统计与决策, 2016(02): 80-82.