



**本科毕业论文**

|  |
| --- |
| **集成预测方法及其在农产品期货市场预测中的应用** |
|  |

**何钏**

**201330810405**

|  |  |
| --- | --- |
| 指导教师 | **张大斌 教授** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学院名称 |  | **数学与信息学院** | 专业名称 |  | **信息管理与信息系统** |
| 论文提交日期 |  | 2017年4月 日 | 论文答辩日期 |  | 2017年5月 日 |

摘 要

随着大数据时代的到来，预测在现代社会中的作用不断凸显。使用现有的单一预测方法只能从某一角度提供有效信息，容易忽略从其他角度获取信息。集成预测的出现了填补这一不足，它能有效整合各单一预测方法，综合利用不同角度的预测信息，以获得更精确的预测效果。

本文重点研究集成预测算法的原理、实现方法以及其在农产品期货的应用。首先简要介绍单一预测中的三大类预测模型，包括时间序列、因果预测、神经网络，重点分析ARIMA、VAR和BP神经网络预测的基本原理和特点。提出了基于融合观察法、排除法、构造排除法来确定ARIMA参数的改进方法。其次针对线性集成的权重确定问题，采用基于等权、误差绝对值和误差平方和三种方式来确定权重，并比较其性能。再次，在非线性集成方面，分析了基于BP神经网络和支持向量机的集成方法。最后，以郑州商品交易所的棉花期货数据为例，借助Eviews和Matlab综合上述预测方法预测期货结算价格。结果表明集成预测在稳定性和预测精度上明显优于单一预测。

关键词 集成预测 单一预测 期货价格 神经网络 支持向量机

**Integrated forecasting method and its application in the prediction of agricultural futures market**

He Chuan

(College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China)

**Abstract:** With the arrival of the era of big data, the role of prediction in the modern society has become increasingly prominent. The existing single prediction method can only provide effective information from a certain angle, and it is easy to ignore the information obtained from other angles. The integration prediction has filled up this problem, it can effectively integrate each single prediction method, and make use of the forecast information of different angles to obtain more accurate prediction results.

In this paper, we focus on the principle and implementation of ensemble prediction algorithm and its application in agricultural futures. The first time in the series focuses on single prediction, causal prediction, neural network three types, including ARIMA, VAR and BP neural network is emphasized, and the latest research on improved methods of fusion method, observation method, rule out structural exclusion method for determining the parameters of ARIMA. Secondly, in-depth study of integration of linear and nonlinear, linear integration focuses on the right, based on the absolute value of error and error square three weighting method based on nonlinear integrated vector machine based on the two kinds of common integration methods of BP neural network, based on support. Finally, taking the cotton futures data of Zhengzhou commodity exchange as an example, this paper forecasts the futures settlement price with the help of Eviews and Matlab. The results show that the ensemble prediction is more prominent than the single prediction in the stability and prediction accuracy.

**Key words:** integrated prediction single prediction futures prices neural network support vector machine

**目 录**

[1 绪论 1](#_Toc480295294)

[1.1 研究背景 1](#_Toc480295295)

[1.2 研究现状 1](#_Toc480295296)

[1.2.1 国外研究现状 1](#_Toc480295297)

[1.2.2 国内研究现状 2](#_Toc480295298)

[1.3 研究内容和研究意义 3](#_Toc480295299)

[2 预测方法介绍 4](#_Toc480295300)

[2.1 ARIMA模型 5](#_Toc480295301)

[2.1.1 ARIMA基本思想 5](#_Toc480295302)

[2.1.2 ARIMA模型参数选择 6](#_Toc480295303)

[2.2 VAR模型 8](#_Toc480295304)

[2.3 BP神经网络 9](#_Toc480295305)

[3 集成预测方法 10](#_Toc480295306)

[3.1 集成预测方法简介 10](#_Toc480295307)

[3.2 线性集成预测 10](#_Toc480295308)

[3.2.1 等权平均集成预测 11](#_Toc480295309)

[3.2.2 变权集成预测 11](#_Toc480295310)

[3.2.3 线性集成预测的实例分析 12](#_Toc480295311)

[3.3 非线性集成预测 16](#_Toc480295312)

[3.3.1 基于BP神经网络的非线性集成预测 16](#_Toc480295313)

[3.3.2 基于支持向量机的集成预测 17](#_Toc480295314)

[3.3.3 非线性集成预测的实例分析 18](#_Toc480295315)

[4 集成预测方法在农产品期货中的应用 20](#_Toc480295316)

[4.1 数据说明 20](#_Toc480295317)

[4.1.1 农产品期货市场研究现状 20](#_Toc480295318)

[4.1.2 数据简介及预处理 21](#_Toc480295319)

[4.2 单一模型进行棉花期货价格预测 22](#_Toc480295320)

[4.2.1 ARIMA模型进行棉花价格预测 22](#_Toc480295321)

[4.2.2 VAR模型进行棉花价格预测 28](#_Toc480295322)

[4.2.3 BP神经网络进行棉花价格预测 32](#_Toc480295323)

[4.3 集成模型进行棉花期货价格预测 35](#_Toc480295324)

[4.3.1 线性集成进行棉花价格预测 36](#_Toc480295325)

[4.3.2 非线性集成进行棉花价格预测 38](#_Toc480295326)

[4.3.3 实例预测结果分析与结论 41](#_Toc480295327)

[5 总结与展望 45](#_Toc480295328)

[5.1 总结 45](#_Toc480295329)

[5.2 展望 46](#_Toc480295330)

[参考文献 47](#_Toc480295331)

[附录 49](#_Toc480295332)

[附录1 BP神经网络代码 49](#_Toc480295333)

[附录2 支持向量机代码 50](#_Toc480295334)

[致谢 51](#_Toc480295335)

华南农业大学本科生毕业论文（或设计）成绩评定表

1 绪论

1.1 研究背景

预测作为一种社会实践的产物，已经历上千年的历史变迁，它是为适应社会经济的发展和管理的需求而产生和发展起来的（陈华友，2008），从开始出现人类古代文明的那一刻起，预测活动也开始出现，并且成为人们日常生活的一部分。随着时代不断发展，科学技术让愚昧和无知逐渐远离我们的生活，预测也开始不再依赖迷信预测和经验预测，逐渐拥有理论的核心，并开始拥有专业的预测方式、应用模型。发展至今，预测已被各个学科领域所接受，担负起分析不同的社会现象和经济活动的重任。

目前，预测方法的种类已多达近百种，加之新方法不断被提出，旧方法不断被改进，这让预测者在选择时需要反复比对。同时不同的预测方法由于原理不同，呈现出来的预测信息也不尽相同，使用者很难权衡优劣，进行取舍。还可能因为某种方法的预测结果存在较大的误差或稳定性不强而被直接放弃，造成其他有用信息不能得到保存。为避免这种情况，J.M.Bates和G.W.J.Granger首次在《运筹学季刊》上提出结合不同预测方法特征进行联合预测来构建预测模型（J.M.Bates,G.W.J.Granger，1969），这种方式能够全面保留各种模型的有效信息，后被称为集成预测。

集成预测方法既保留传统，又进行创新，它用组合的方式将各个单一模型的结果进行整合。由于综合多种模型并扬长避短，因此其结果准确度通常会高于单一方法。而且随着单一预测方法不断被改进，单一预测结果变得更为精确，集成预测的精准度也得以不断提高，集成预测已成为预测中不可或缺的组成部分。

1.2 研究现状

1.2.1 国外研究现状

国外其实在J.M.Bates和G.W.J.Granger之前就开始使用集成预测。早在19世纪60年代， Schmitt就将两个预测方法的结果进行简单线性集成，预测三十多个城镇中的人口分布情况。只是当时发表的期刊名气较低，并引起社会的关注。到了1969年，J.M.Bates和G.W.J.Granger通过在权威期刊上发表文章，才使集成预测方法被广大的专家学者所知道，并开始了它的生命历程。

从1970年到1980年，有关学者对集成预测更多是停留在理论上的探索，使用的集成也是简单地对多个模型的结果进行平均。后来，Makridakis和Winkler不满足于简单集成，深入研究影响集成效果的深层原因。他们用1001组数据多次进行实验比对，惊奇地发现单一模型的个数会对最终集成预测的结果产生影响，而且随着单一模型数量的增多，预测结果会更加准确（Makridakis,Winkler，1983）。这一发现让有关学者开始不再局限于原有思维，开始转向对集成预测中单一预测方法的研究。Diebold在前人的基础上，详细地分析了各个单一预测方法之间的关联关系和它们对最终集成效果的影响（Diebold，1988）。

1989年，《Journal of Forecasting》充分利用其在预测领域的权威地位，邀请众多该领域的专家一同撰写了多篇与集成预测的相关文章（Anandalingam，1989；Granger，1989；Guerard，1989），并将这些文章制作成合集进行发售。从此集成预测搭上了时代的快车，快速地向前发展。1990年，Diebold根据Winkler和Makridakis发现不同单一预测模型在不同的数据源预测效果不尽相同这一结论，提出要对单一预测方法进行初步筛选，再用贝叶斯法确定权重进行集成，取得了接近以往进行简单平均集成的结果（Diebold，1990）。打破在预测方面的思想禁锢，使集成预测不再局限于简单平均，集成预测进入高速发展阶段。

进入21世纪后，集成预测方法不断走向成熟。Gordon和Michael通过对三种时间序列进行集成预测，发现其中一种预测方法在拟合时的表现非常突出，而另一种在则在预测结果上非常接近实际值（Gordon,Michael，2003），进而让众多的学者开始关注到不同预测方法的拟合和预测效果适用性这一层面的讨论。在这阶段中对集成预测的实验更多是单一因素分析，缺少多因素约束下的集成预测分析。David E.Rapach和Jack K.Strauss使用不同的预测模型结合多个影响房价的有关因素进行分析，通过几种不同的集成方式建立了回归集成预测方法，大大提高了预测的精度（David E.Rapach, Jack K.Strauss，2009）。

1.2.2 国内研究现状

我国在前期受国情影响，与其他发达国家相比在集成预测上投入的精力和资金相对较少。但从改革开放后，随着经济不断发展，国家对于科学研究越发重视，集成预测也在我国的大地上开始快速发展。目前，我国在集成预测的研究上获得了很大的成就，唐小我教授、陈友华教授、曹长修教授等人通过不断地深入研究有关文献，先后发表了一系列的关于集成预测的研究论文。我国的《预测》从1984年起，每年都会刊登有关集成预测的相关文献，这为集成预测的进一步发展做了很好的铺垫。

在集成预测的发展中，主要是以先行集成预测和非线性集成预测作为两大研究方向，下面就这两个进行展开。

在线性集成预测方面，以孔庆凯提出直接平均法集成拉开序幕（孔庆凯，1985），并带动有关学者研究直接平均法集成适用范围的热情。唐小我借助误差平方和来确定线性集成的最优加权系数（唐小我，1991）。唐小我深入研究变权重集成，并提出权重和应在任一时刻点都为1（唐小我，1993）。邓雪利用矩阵理论划出误差平方和的取值范围，提高了简单平均集成的预测准确性（邓雪，2001）。陈雁以Kuhn-Tucker为基础条件用Wolfe方法研究非负权重最优模型集成预测方法解法（陈雁，2007），打破了只能进行线性集成预测的研究成果。高尚等用最大相对误差最小、相对误差方法和最小和相对误差和最小作为标准，提出多个线性集成模型，并证明这些模型的优越性和准确性（高尚，张绍彪，梅亮，2008）。

在非线性集成方面，文新辉、牛明洁在国内首次发表了有关非线性集成模型，借助BP神经网络集成，取得了优于线性集成的预测结果（文新辉,牛明洁，1994）。董景荣提出基于小波网络的非线性集成方法，该方法解决了以往在对非平稳的时间序列使用线性集成时所造成的预测效果不佳的困境（董景荣，2000）。李元诚、李波等人提出基于小波支持向量机的非线性集成并用于仿真实验，发现该模型比神经网络集成更为优异（李元诚，李波，方廷健，2004），为非线性集成模型的选取提供了新思路。耿建军先用二次规划确定权重，再用广义神经网络训练进一步确定权重，有效提高预测精度，为变权集成预测开创了新的方法（耿建军，2011）。李钦铭和芮兰兰提出了基于K-means进行序列分隔后采用对每个序列分别建模，最终预测误差只有BP神经网络的1/10，为非线性集成预测提供新思路（李钦铭，芮兰兰，2016）。

1.3 研究内容和研究意义

本文主要对单一预测和集成预测进行探索，再将其用农产品期货市场价格为例进行操作，展现集成预测比单一预测的突出表现。

首先，在绪论部分简单介绍了预测的基本概念和发展过程，包括对单一预测方法所存在一些不足的说明、集成预测方法提出的背景以及对国内外集成预测的研究过程和现状。其次，简单介绍单一预测的时间序列、因果预测、神经网络三大类型，并以其中最具代表性的ARIMA模型、VAR模型和BP神经网络为例进行介绍。再次，分别介绍线性集成与非线性集成，说明建模思路，列举典型的集成方式进深入阐述。在线性集成部分分析了等权、误差平方和、误差绝对值三种确定权重的方法。在非线性集成部分举例神经网络、支持向量机两种方法。单一模型的确定和集成部分是本文的重点和难点，因为单一预测方法的选择使用会影响集成预测方法的输入样本的准确度，所以不仅要合理选择集成预测方法，也要兼顾单一预测方法自身的特点，因此这部分内容决定最终预测结果的好坏。

最后，利用农产品期货市场价格来进行实际预测，建立ARIMA模型、VAR模型和BP神经网络进行单一预测，再用线性集成和非线性集成的方式对单一预测方法的结果进行集成，再用表格的形式对多种预测结果进行对比分析，进而推出集成预测在预测中的突出表现。

本文的研究意义一方面在于对集成预测方法的进行系统研究，用理论加实践的方式说明集成预测在预测在提高预测结果精度的突出作用，另一方面在于通过对现在人们比较关注的农产品期货市场进行预测，运用实际的例子表明集成预测在实际生活中的重要作用，通过抛砖引玉的方式为集成预测在其他领域的运用作铺垫。

2 预测方法介绍

现有的预测方法由于它们本身具有的不同特性决定了它们的适用范围各不相同。由于预测方法非常多，不能对所有的预测方法进行一一列举介绍，所以本文通过对现有的预测方法大类进行简单介绍，并对其中最具代表性的模型深入阐述。在大类上，预测可分为定性预测和定量预测。

定性预测是根据使用者的前期经验或是相关学者的评估意见进行判断的方法，它借助过往的经验预测未来的走向，是基于人为主观判断的，因此对使用者的经验要求相对较高。同时，在这过程中难以采用具体的量化指标来科学解释预测结果，所以在生活中较少使用。只有在没有相关数据支持的情况下，才会使用这类方法。常用的定性预测方法包括直观法、专家判断法、德尔菲法等。由于在定性预测上不能很直观地判断模型好坏，所以本文不对此部分进行展开实践。

定量预测是指用往期的数据作为输入，根据数据类型选择合适的公式或算法进行建模，通过不断调整参数设置来提高预测精度。它主要分为以传统的计量经济学为基础的统计方法和以人工智能为基础的机器学习预测方法。在计量经济学中，包括时间序列法和因果法，机器学习中神经网络最为常用。本文将通过对时间序列法、因果法、神经网络法中的最具代表性的模型进行介绍来加深对定量预测的讲解。其中，时间序列法以ARIMA模型作为例子、因果法以VAR模型作为例子、神经网络以BP神经网络作为例子进行介绍。

2.1 ARIMA模型

2.1.1 ARIMA基本思想

ARMA模型（自回归滑动平均模型）是时间序列法最常用的模型，它是由AR模型（自回归模型）和MA模型（移动平均模型）组合而成的，均用于平稳时间序列。对非平稳序列，先对数据差分变为平稳序列，再用ARMA模型，这就是ARIMA模型（综合自回归移动平均模型）。

ARIMA模型发展较早，因此有丰富的理论知识基础。其能结合历史的数据发掘规律，得到较为准确的预测结果，加上其相对简便的操作方法，使其成为众多进行时间序列预测的使用者的常用模型。

（1）AR模型（自回归模型）是对时间序列的滞后项进行加权并添加随机干扰项结合而成，其公式为：

(1)

其中， 是常数；是AR；是AR；是AR模型残差。检验残差是否通过白噪声检验，判断公式为：。

（2） MA模型（移动平均模型）是对白噪声序列进行加权平均，其公式为：

(2)

其中， 是常数；是MA；是MA；是MA模型残差。检验残差是否通过白噪声序列，判断公式为：。

（3）ARMA(p,q)模型是由AR模型和MA模型组合而成，当ARMA模型的p等于0时，则模型变为MA模型，当q=0时，模型变为AR(p)。其具体的表示形式为：

(3)

其中， 是常数；是AR；是AR；是AR模型残差；是MA；是MA；是MA。

（4）ARIMA(p,d,q)模型是基于差分和ARMA(p,q)模型组合而成，它依靠滞后期数和干扰项相结合来确定模型。对非平稳的时间序列，先进行差分，直至序列平稳时结束，此时差分的次数记为d，再用常规的ARMA(p,q)模型进行预测。在判断序列是否平稳时，常用的方法包括Augumented Dickey-Fuller检验、Phillips-Perrson检验、Ng person Tests检验。

2.1.2 ARIMA模型参数选择

由ARIMA(p,d,q)模型的简介中可以看出，模型中p值和q值的选择是否准确影响着预测效果好坏，p值和q值越符合数据，则模型后期的预测结果就越准确。为了提高ARIMA模型的预测准确度，本文对p值和q值确定方法进行研究。列举出现在最常用的p值和q值确定的方法如表1。

表1 常见p值和q值确定方法

|  |  |  |  |
| --- | --- | --- | --- |
|  | 观察法 | 排除法 | 构造排除法 |
| 操作方式 | 对在前期绘制出的自相关图和偏自相关图，通过目测的方式来确定p值和q值。 | 先初步确定p值和q值，再排除不显著系数，反复操作，直至留下通过显著性检验的系数。 | 先初步确定p值和q值，再对S C值或AIC值小的组合检测系数显著性，如通过检验，则确定p值和q值，否则再向上检验第二小的组合的显著性，反复操作直至检验通过，此时的p值和 q值即为最终的结果。 |
| 优点 | 操作方便，无需多次重复。 | 对确定的p值和q值构成的组合多次检验，排除了不合适的系数。 | 排除了人为的主观性，尝试了多种不同组合 |
| 缺点 | 确定过程过于依赖人为确定，结果过于主观。 | 确定过程只考虑一种组合，没有筛选出最优的p值和q值的组合。 | 没有充分利用AIC值和SC值 |

以上三种为最常见确定p值和q值的方法。笔者参阅了多份文献后，最终确定以陈林提出的结合以上三种确定p值和q值的方法作为本文的实验方式（陈林，2010）。下面对该方法进行详细介绍。

首先使用观察法初步确定p值和q值。第二，根据p值和q值列出所有可能的组合绘制成表格，本文以p=2，q=2进行举例，绘制图表如表2。

第三，根据绘制出的表格逐一地对所有的组合进行系数的显著性检验，排除所有系数不通过检验的组合。

表2 ARIMA中p=2和q=2的组合过程

|  |  |  |
| --- | --- | --- |
| AR模型 | MA模型 | 组合 |
| AR(2) AR(1) | MA(2) MA(1) | {AR(1)}; {AR(2)}; {MA(1)}; {MA(2)}; {AR(2),MA(1)}; {AR(2),MA(2)};  {AR(1),MA(1)}; {AR(1),MA(2)}; {AR(1),AR(2),MA(1)}; {AR(1),AR(2),MA(2)}; {AR(1),MA(1),MA(2)}; {AR(2),MA(1),MA(2)}; {AR(1),AR(2),MA(1),MA(2)} |

第四，对通过显著性检验的组合分别计算它们的AIC值和SC值，选择出AIC值和SC值都最小的组合，如果AIC值和SC值分别在两个组合中最小时，则再根据这两个组合对训练集拟合得出的RMSE值，选择最小的作为最优组合。此时的p值和q值则结合了原始的三种常规的确定方法，其预测结果有较好的性能表现。

2.2 VAR模型

VAR模型全称为向量自回归模型，该模型通过对统计学中的数据进行相关特性的描述，挖掘所有数据间不同属性之间可能存在的关联关系，再确定各个属性的相互影响的方法，再将她们确定下来形成一个整体作为预测指定数据的输入集，并结合这些输入集中的自身历史变化情况来确定它们所能影响的滞后期数和影响方式，以此来预测建模。这个模型已经在很多的领域上受到了重用，各类的专家学者也开始对其在预测能力上有更多的肯定。

VAR模型的p阶表达形式为：

(4)

其中，是n维的属性向量；是n\*n维的向量矩阵；是m维外生量；是n\*m维矩阵；是k维的误差向量。

在VAR模型中，对于滞后期数p的确定是整个模型中非常重要的一个环节，一个错误的影响期数哪怕进行再近似的拟合和预测，也会存在极大的误差，所以滞后期数的确定显得尤为重要。在对其进行确定上，有关的学者提出了多种的确定方法，其中较为常用的为AIC法、SC法以及似然检验法。这些方法都是通过各种准则来确定最优的期数，进而来保证能够使最终的残差不存在自相关性。现阶段，得益于分析软件的不断完善，现在在进行滞后期数确定时，可以通过多种准则共同判断，选择出最优的滞后阶数。

在VAR建模的过程中，会涉及多个影响因素，因为在主观层面上很难判断这些影响因素是否与最终的预测目标有关，所以通过科学的方式确定各个因素的关系尤为重要。对因素进行相关性分析和格兰杰因果关系检验是最常用的检验方式，由相关性分析发现各个因素之间是否存在关联关系，排除没有关联的因素。再通过格兰杰因果关系检验可以确定这些有关的因素是否是最终预测目标的前置原因，如果检验结果是肯定的，则可以直接进行VAR建模。

2.3 BP神经网络

人工神经网络是近几十年来在机器学习中的一个重点，由于它不受到各类固定式子的限制，能够以更加高效的方式不断地逼近最终的实际值，所以成为了各类学者在预测时优先选择的预测模型。神经网络是通过模拟人类的神经系统中的神经元的工作模式，由网络自动对数据进行学习总结得出不局限于特定公式的预测结果的预测模型。BP神经网络则是在人工神经网络中重要的一个组成部分，也是神经网络中应用最为广泛的模型。

BP神经网络是由三个部分组成，包括最初的输入层、中间部分的若干个的隐藏层，以及在最末端的输出层。在这些层级之间，只有相邻的两个层级之间会有所连接，其他层级不会有任何关联，同时同一层的中神经元也不会有所关联，它们各自独立，互不影响。下图为最常规的三层的BP神经网络结构图：

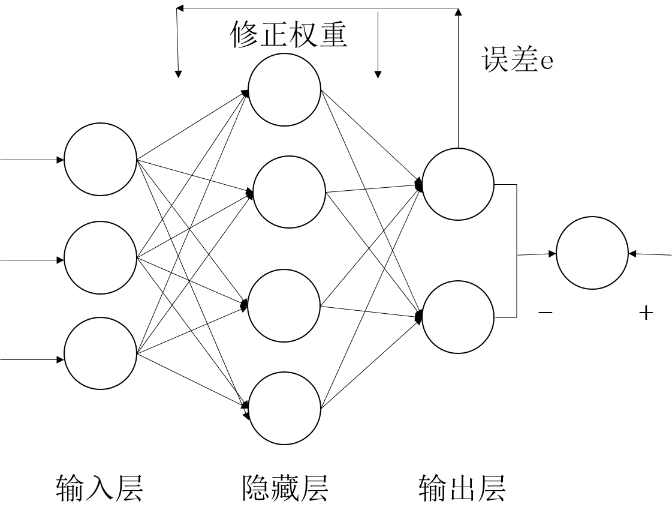


图1 三层BP神经网络结构图

BP神经网络的BP是指反向传播，但反向传播的是误差，其实其正真的实质是通过误差逆向传播算法来训练的前馈神经网络。其通过在输入层处输入数值，再由一开始设定的隐藏层层数和神经元个数来形成相应的映射关系。虽然没有直接的公式进行计算，但是其在一开始时会对各个神经元赋予初始的阀值和权重，再通过这些初始数值经过多层隐藏层进行训练，得出输出的结果，再将输出结果与真实值进行比对，确定误差大小，按照刚刚各个神经元赋予的权重再次将误差分配回去，并根据这个返回的过程调整权重，按照这个流程多次重复，直到将此神经网络训练到能够达到预期的效果才停止下来。同时在这个过程中，训练出来的结果的误差会慢慢下降，这个过程也叫梯度下降法。

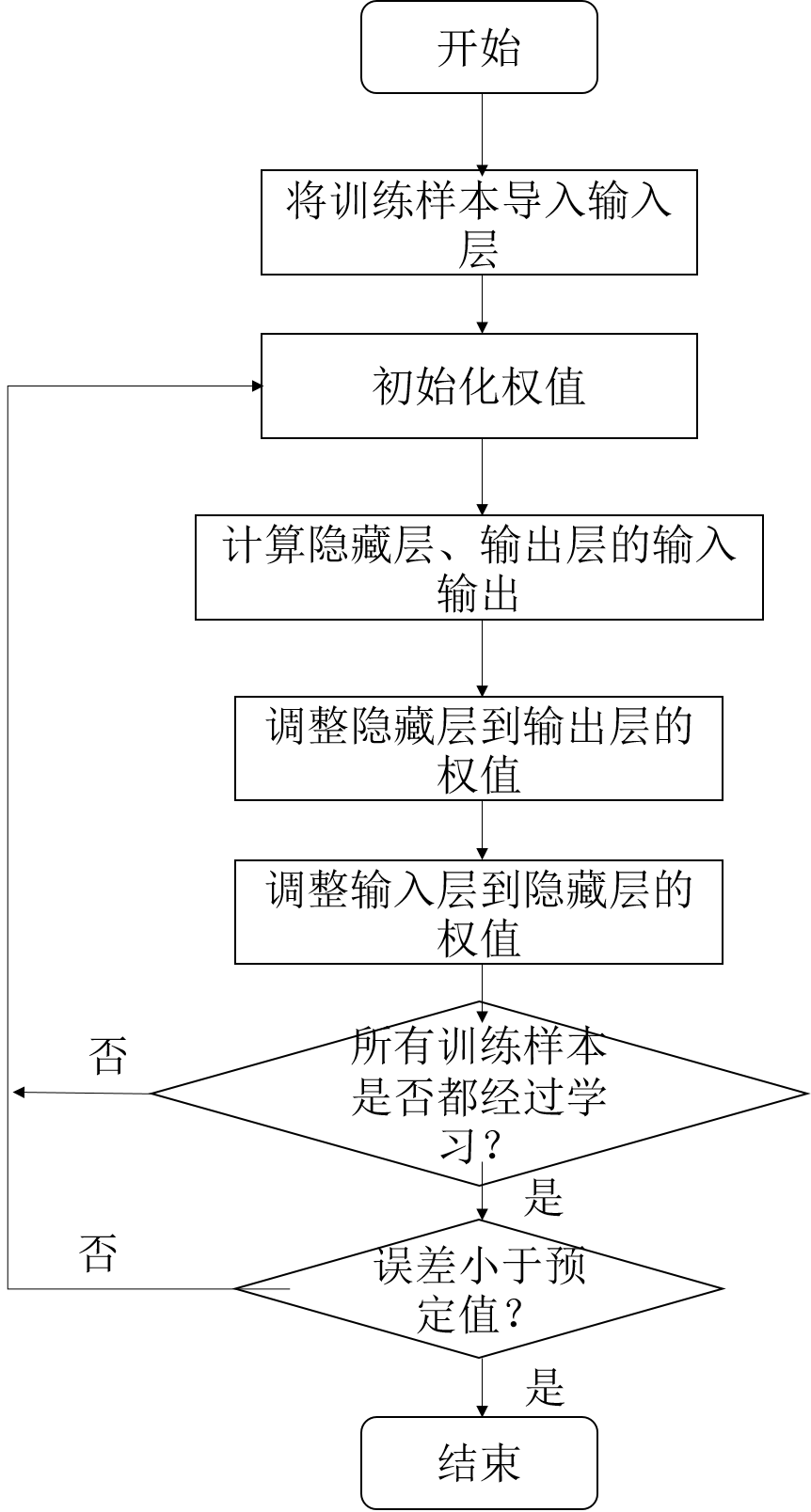


图2 BP神经网络训练过程

3 集成预测方法

3.1 集成预测方法简介

现阶段已经有很多的各类单一预测模型被提出，针对相同的数据，采用不同的模型、设置不同的参数就会产生不同的预测效果。每种单一预测方法都在预测结果的呈现上都有其各自不同的特点，如果只是采取其中的一种方法，则可能会失去其他方法所能提供的信息，所以通过将不同的单一模型组合起来，通过调整它们在组合后的新模型中的比重，来形成新的预测方法就被称之为集成预测方法。它最大的优点就是它填补了单一预测方法容易受到方法设置上的随机因素的影响，通过整合各个模型的有效信息，进而提供更加全面准确的预测结果。

3.2 线性集成预测

线性集成预测在众多的文献中都被称为组合预测，它是通过对各种单一预测模型的预测结果进行分析，比如结合它们的预测值与真实值之间的误差，借助不同的计算方法、不同的侧重来确定每个模型应有的权重，进而提升最终的预测结果的准确性。

对于线性集成预测，其一般的表示形式为：

(5)

其中，是第i个模型在t期的预测值；是第i个模型在线性集成预测方法中的权重。对于权重的系数，应该满足：

(6)

并且，在实际中还要模型的权系数还要求为非负即。

在线性集成预测方法中，确定权重系数的方式有很多种，比如等权平均法、方差倒数法、残差倒数法，本文将简单地对其中的常见的确定方式进行介绍。

3.2.1 等权平均集成预测

等权平均集成预测法是最先被提出的一种集成预测方法，它最大的优点就是不需要分析各个单一预测方法的预测效果如何，也不需要进行复杂的计算赋予它们权重，直接通过平均化的方式，给予每个单一方法同样的比重后再进行相加。

等权平均集成预测法的表示形式如下：

(7)

其中，是第i个模型的预测值。

虽然此方法是最简单最快捷的一种，可能存在不能合理利用各个单一模型优点的问题，但是它却是在不知道各个模型的预测准确度、现实中实际值的情况下，可以采用的最为安全、保险的一种集成方法。

3.2.2 变权集成预测

变权集成预测是线性集成预测中最为常用的组合集成方式，它的基本原理就是通过对计算出各个单一预测方法的预测值与实际值之间的误差，对这些误差通过某种计算累计形成各个模型所产生的总误差，对于误差较小的模型赋予较大的权重，对误差较大的模型赋予较大的权重，进而使得最终集成后的预测值更加接近于真实值，减小误差。本文将列举出基于误差绝对值和基于误差平方和这两个变权集成方法来进行介绍。

（1）基于误差平方和的变权集成方法

通过计算各个模型的误差绝对值之和，再给误差和最小的模型赋予最大的权重的方式来进行集成的方式，是变权集成预测中常用的方法之一。其表示形式为：

(8)

(9)

其中，是第j个单一预测模型的误差绝对值之和；是第t期实际观察值；是第t期的第j种方法的预测值。对于权重的系数，应该满足：

(10)

并且，在实际中还要模型的权系数还要求为非负即。

（2）基于误差平方和的变权集成方法

通过计算各个模型的误差平方和，再给平方和最小的模型赋予最大的权重的方式来进行集成的方式，是变权集成预测中常用的方法之一。其表示形式为：

(11)

(12)

其中，是第j个单一预测模型的误差平方和；是第t期实际观察值；是第t期的第j种方法的预测值。对于权重的系数，应该满足：

(13)

并且，在实际中还要模型的权系数还要求为非负即。

3.2.3 线性集成预测的实例分析

在这一小节中，主要是针对以上提及的几种线性集成预测算法进行简单的进行运用，并针对运用的结果进行简单的分析。在引入实例分析之前，先对目前较为常见的判断模型预测效果好坏的指数进行简单的介绍。

表3 常见判别公式

|  |  |
| --- | --- |
| 判断方法 | 公式 |
| 残差平方和 |  |
| 平均绝对误差 |  |
| 平均绝对百分比误差 |  |
| 均方误差 |  |
| 均方绝对百分比误差 |  |

本文在对各种实例进行分析时，就以表格中所列举的残差平方和、平均绝对误差和平均绝对百分比误差作为判断模型预测好坏的依据。现使用2015年大豆期货数据在采用专家模型和指数平滑模型后得出来的234个拟合数据、10个预测数据和实际数据来进行线性集成预测的实例分析，下图为经过整理汇总后的预测数据：



图3 2015年大豆期货原预测数据汇总

为了更直观地比较出这两种模型在预测上效果的好坏，本文先对所有的模型计算出每个年份的误差绝对值，整理后如下图：

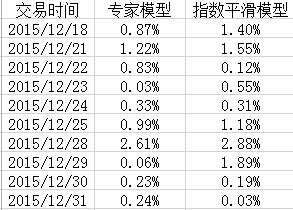


图4 2015年大豆预测数据误差百分比

由此可以直观地发现这两个模型的误差相对较大，而且不同年份之间的变化波动也较为明显。对此，开始引入本章中提到的三种线性集成预测方法。

对于等权平均集成预测，由于不需要分析各类单一预测模型，且实例中的模型共有3种，所以每个模型的权重都为0.5。此时的线性集成表达式为：

(14)

其中，是专家模型在第t期的预测结果；是指数平滑模型在第t期的预测结果。

对于基于误差绝对值的变权集成预测，先要将各个模型拟合的误差绝对值进行累加取平均后，再取倒数来分配权重，具体公式参考上文。通过累加计算可知，专家模型的平均误差绝对值为27.1990，指数平滑为37.0637。根据上文的公式可求出专家模型的权重为0.5768，类似地，指数平滑为0.4232。此时根据误差绝对值的线性集成预测的表达式为：

(15)

其中，是专家模型在第t期的预测结果；是指数平滑模型在第t期的预测结果。

对于基于误差平方和的变权集成预测，先分别将各个模型拟合的误差平方和相加得出其总误差平方和，再取倒数分配权重。公式参考上文，经过计算可知，专家模型的误差平方和为14505.26，指数平滑为24052.65。根据取倒数分配权重的方式，专家模型的权重为0.6238，类似地，指数平滑为0.3762。此时根据误差平方和的线性集成预测的表达式为：

(16)

其中，是专家模型在第t期的预测结果；是指数平滑模型在第t期的预测结果。

分别将这些线性集成方法得出的每一年预测值及误差整理得到下表。

表4 大豆期货线性集成预测值及误差

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 交易日期 | 等权平均集成预测 | | 基于误差绝对值变权集成预测 | | 基于误差平方和变权集成预测 | |
| 预测值 | 误差 | 预测值 | 误差 | 预测值 | 误差 |
| 2015/12/18 | 3683.62 | 1.14% | 3685.13 | 1.10% | 3686.06 | 0.32% |
| 2015/12/21 | 3697.11 | 1.38% | 3698.08 | 1.36% | 3698.67 | 8.19% |
| 2015/12/22 | 3719.07 | 0.35% | 3721.78 | 0.43% | 3723.44 | 6.47% |
| 2015/12/23 | 3723.89 | 0.29% | 3722.41 | 0.25% | 3721.51 | 5.81% |
| 2015/12/24 | 3712.95 | 0.32% | 3713.01 | 0.32% | 3713.05 | 3.33% |
| 2015/12/25 | 3707.86 | 1.09% | 3707.31 | 1.07% | 3706.98 | 6.70% |
| 2015/12/28 | 3697.69 | 2.74% | 3696.94 | 2.72% | 3696.48 | 4.34% |
| 2015/12/29 | 3657.45 | 0.98% | 3652.36 | 0.84% | 3649.24 | 3.59% |
| 2015/12/30 | 3623.39 | 0.21% | 3623.27 | 0.21% | 3623.20 | 2.03% |
| 2015/12/31 | 3624.96 | 0.14% | 3625.54 | 0.15% | 3625.90 | 0.09% |

可以明显看出这几种线性集成的方法已经将原先在单一预测模型中的前两种模型所带有的稳定性不足的缺点进行了一定的改善。为了能够进行一步的对比，本文通过将各个单一预测模型和三种线性集成方法的残差平方和、平均绝对误差和平均绝对百分比误差绘制在一张表格中进行再次比较。

表5 大豆期货线性集成预测效果评估

|  |  |  |  |
| --- | --- | --- | --- |
|  | 残差平方和 | 平均绝对误差 | 平均绝对百分比 |
| 专家模型 | 14505.26 | 27.20 | 0.74% |
| 指数平滑 | 24052.65 | 37.06 | 1.01% |
| 等权平均集成预测 | 17588.37 | 31.67 | 0.86% |
| 基于误差绝对值变权集成预测 | 16895.42 | 30.99 | 0.85% |
| 基于误差平方和变权集成预测 | 16510 | 30.57 | 0.83% |

从这张表格中，可以很直观地看出在这三种检验中三种集成方法都有比较良好的表现，特别是基于误差平方和的集成预测方法，但此时可以发现虽然三种线性集成的方法都给数据带来了预测的稳定性，但是在残差平方和等这三种结果评价的结果上都是没有专家模型单个的效果好，这是因为这些模型都是整合了两种模型的结果来分配权重，同时分配权重的时候只能是正数，所以效果只能够保持在相对较优的状态下。但总体来说，线性集成预测预测结果会比只用一种预测方法来得稳定，但有时可能也会存在不如单一预测效果好的现象。同时虽然在这个实例中基于误差平方和的集成预测方法的表现最好，但是这并不是意味着能够对所有的实例数据都适用，不同的集成方法在不同的数据源会有不同表现。不过从众多的文献材料中以及本文的实例验证，变权集成预测的预测效果往往会比等权集成预测的效果更好。所以在进行线性集成预测中，最好的选择是将各种集成方法都进行实验，选取最优的作为该数据的指定模型。

3.3 非线性集成预测

对于线性集成预测的方式，就如同上一部分所讲述的，它们能够直接明确地通过公式进行展示，在向他人进行解释说明时也能够以非常直观的方式进行呈现。但却存在不能够结合不同的时间点进行分析，不够灵活，同时在权重确定方式上也容易受到人为选择的影响，比较难产生较优的结果的局限性。所以有关专家学者就提出了通过使用非线性的方式进行集成，比如基于神经网络、基于支持向量机等。

3.3.1 基于BP神经网络的非线性集成预测

基于BP神经网络的非线性集成预测其实类似于在前文提到的用BP神经网络进行预测，只是在输入的变量上换成了多个单一预测模型的预测值，作为为神经网络的输入样本，而实际的真实值作为最终的输出样本。通过使用BP神经网络的误差反向反馈的方式，让误差分配到每个指定的神经元，进而来改变每个神经元分配到的权重，最终输出经过学习训练后的输出样本，作为最终非线性集成的结果。

但是在BP神经网络在进行集成预测时，最大的问题就是要如何确定隐藏层的个数，以及每个隐藏层中神经元的个数。经过众多的学者研究发现，只要能够合理地设置网络结构、神经元节点数以及每个节点分配合适的权重，其实最为基础的三层神经网络也是可以得到很好的预测拟合效果。虽然借助增加隐藏层层数的方式能够更好地提高预测的精度，降低误差率，但是由于神经网络本身就是非常复杂的过程，过多的层数只会让网络变得更加复杂，增加了网络学习的时间。所以本文在进行有关BP神经网络预测的实验，不管是单一预测还是非线性集成都会采用三层的神经网络结构，通过不断改变神经元的个数，来调试出最合适的神经元个数。常用的确定神经元节点数的经验公式如下。

(17)

其中，m是隐藏层的节点数；n是输入层的节点数；l是输出层的节点数；是1-10之间的常数。

3.3.2 基于支持向量机的集成预测

支持向量机是近年来在机器学习界中众多专家学者研究的热门模型，它能够同时满足置信范围最小和经验风险最小这两个要求，并且它不像传统的机器学习如神经网络之类的模型有着需要依赖多维数来提高预测精度的问题，降低了模型的复杂度，因此即拥有了优秀的学习拟合能力，又更便于推广。

支持向量机训练的原理在于它能够将非线性集合通过非线性的函数映射到一个高维度的空间中，并采用满足Mercer条件的内积核函数来进行回归，这避开了由因高维特征空间引起的维数灾难，其函数表达式为：

(18)

\\

而基于支持向量机的非线性集成预测模型是将各个单一预测模型的拟合值和预测值作为支持向量机的输入值，并将与之对应的实际值作为输出值，通过采用大量的实例进行训练，让所有的单一预测模型都能够与最终的实际值形成一种非线性的映射关系【挖掘出多个单一预测值与实际值之间的内在联系】，并且通过不断地训练，使模型达到预期的精度。

而在实际应用过程中，由于支持向量机在寻找最优参数的过程中需要经过非常复杂的调试过程[由于支持向量机训练过程实际上是参数寻优的过程，能否得出最佳的模型参数觉得这模型的预测效果]，于是就出现了网格搜索、遗传算法和粒子群算法等多种寻参方式，通过这些算法的自动寻参，再配合上交叉验证等方式确定最佳参数，就可以极大地提高支持向量机的使用方便性。

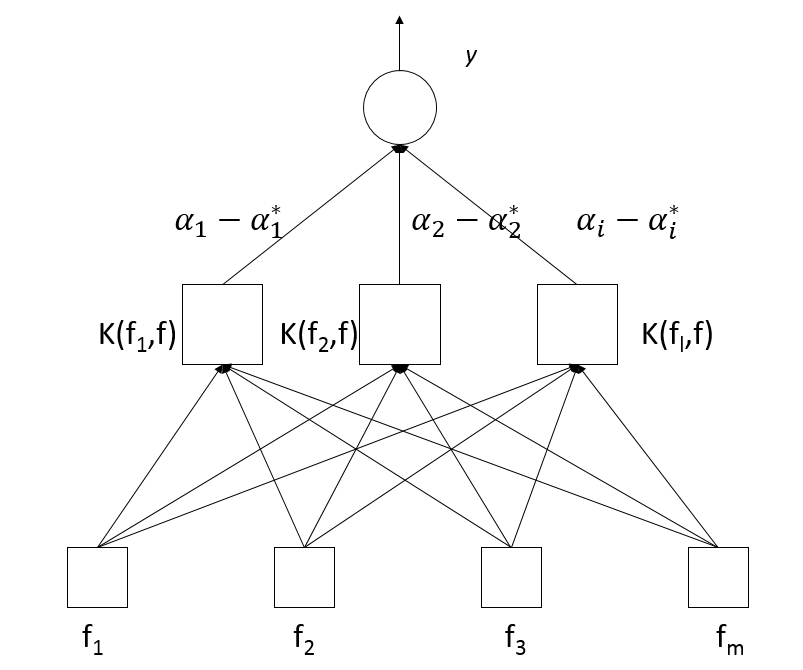


图5 支持向量机结构图

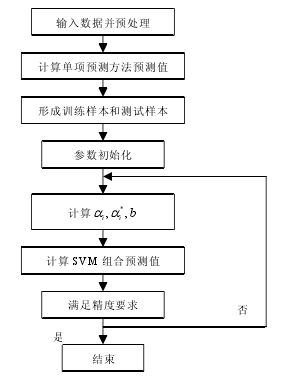


图6 支持向量机预测流程图

3.3.3 非线性集成预测的实例分析

在本小节中，使用的数据实例依旧是大豆期货的专家模型和指数平滑模型的预测结果。由于本节只是简单地说明非线性集成预测在预测效果提升上的表现，所以对于中间进行神经网络和支持向量机的调试方面就省略不讲。

采用神经网络进行集成时，经过不断测试，最佳的模型是三层网络结构，隐藏层中神经元个数为3的BP神经网络模型。在训练神经网络时用234天的拟合数据作为训练集，而10天的预测数据作为测试集。

采用支持向量机进行集成时，经过不断测试，最佳的模型是基于粒子群算法自动寻参。同样地，也就是将前234天的拟合数据作为训练集，后10天的数据作为测试集。

分别将基于神经网络和基于支持向量机的非线性集成得出的每一年预测值及误差整理得到下表：

表6 大豆期货非线性集成预测值及误差

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 交易日期 | 基于神经网络的非线性集成预测 | | 基于支持向量机的非线性集成预测 | |
| 预测值 | 误差 | 预测值 | 误差 |
| 2015/12/18 | 3695.88 | 0.81% | 3700.14 | 0.69% |
| 2015/12/21 | 3704.06 | 1.20% | 3707.23 | 1.11% |
| 2015/12/22 | 3742.78 | 0.99% | 3743.00 | 1.00% |
| 2015/12/23 | 3707.21 | 0.16% | 3709.18 | 0.10% |
| 2015/12/24 | 3710.93 | 0.27% | 3713.25 | 0.33% |
| 2015/12/25 | 3700.71 | 0.89% | 3703.16 | 0.96% |
| 2015/12/28 | 3689.51 | 2.51% | 3692.15 | 2.59% |
| 2015/12/29 | 3624.89 | 0.08% | 3617.63 | 0.12% |
| 2015/12/30 | 3626.17 | 0.13% | 3632.91 | 0.05% |
| 2015/12/31 | 3632.36 | 0.34% | 3640.49 | 0.57% |

可以明显看出这两种非线性集成的方法已经将原先在单一预测模型中的前两种模型所带有的稳定性不足的缺点进行了一定的改善，同时也提高了预测的准确性。为了能够进行一步的对比，本文通过将各个单一预测模型和非线性集成方法的残差平方和、平均绝对误差和平均绝对百分比误差绘制在一张表格中进行再次比较。

表 7 大豆期货非线性集成预测效果评估

|  |  |  |  |
| --- | --- | --- | --- |
|  | 残差平方和 | 平均绝对误差 | 平均绝对百分比 |
| 专家模型 | 14505.26 | 27.20 | 0.74% |
| 指数平滑 | 24052.65 | 37.06 | 1.01% |
| 基于神经网络的非线性集成预测 | 13859.48 | 27.09 | 0.74% |
| 基于支持向量机的非线性集成预测 | 14304.37 | 27.58 | 0.75% |

从这张表格中，可以很直观地看出在这两种集成方法都有非常好的表现，特别是基于神经网络的非线性集成预测方法，它在这三种评估的结果表现上都优于单一预测方法中的任一种方法，而基于支持向量机的集成方法，虽然还是没有专家模型的表现效果好，但因为它能够对平稳预测数据，同时也提高了预测的准确性，所以还是非常值得使用的模型。总体来说，非线性集成预测预测结果会比只用一种预测方法来得稳定，同时在结果评估上也有优异的表现。值得说明的是，本节中是基于神经网络的集成效果好，但在实际中并不一定，所以要结合真实的数据情况进行使用。

在进行了线性集成与非线性集成后，可以明显地感觉到非线性集成的集成效果最终的预测精度会比线性集成的好，这是由于非线性的集成方法不止局限于特定的公式，而是通过机器的不断学习、不断寻找最优参数的方式进行预测，所以结果会相对较好。但是非集成也有局限性，当训练集的个数太少时，就很容易出现训练效果不佳的情况，所以在数据少的情况下，进行线性集成是个较好的方法。

4 集成预测方法在农产品期货中的应用

为了对集成预测方法在提升预测精度方面的效果，有更加直观的感受，本文通过对目前广受大家关注的农产品期货市场进行一个实际操作，通过对比单一预测方法与集成预测方法在预测结果的误差程度的比较，来加深对集成预测方法的了解，同时也可以将此集成预测方法作为农产品期货价格预测的备用选项。

4.1 数据说明

4.1.1 农产品期货市场研究现状

虽然我国的期货市场起步较晚，但有国外的农产品期货市场作借鉴，对农产品期货市场的研究也是国内的一大热点。目前对于农产品期货市场的预测中，主要采用的研究方法有AIDS模型、线性支出系统模型、线性函数模型、对数线性函数模型、ARIMA模型等，而集成预测方法的使用相对较少，王江、费宇（王江，费宇，2010）运用时间序列ARIMA模型和回归模型集成对上海期货进行预测，褚东（褚东，2012）通过对ARMA(2,2)和GARCH(1,1)模型对大豆期货价格进行预测，都取得了不错的预测效果。

4.1.2 数据简介及预处理

目前我国农产品期货市场主要为上海期货交易所、大连期货交易所和郑州商品交易所，农产品在这三大期货市场中都占有极其重要的位置。本文选择郑州商品交易所（ZCE）2016年主要的交易品种棉花作为集成预测的实验数据。

从官方网站上下载的2016年历史行情数据，包含郑州商品交易所2016年所有交易品种每日行情数据，共32687条，每条数据包含了一系列的交易内容，如今开盘、最高价、最低价、今收盘、成交量、成交额等，如图7。

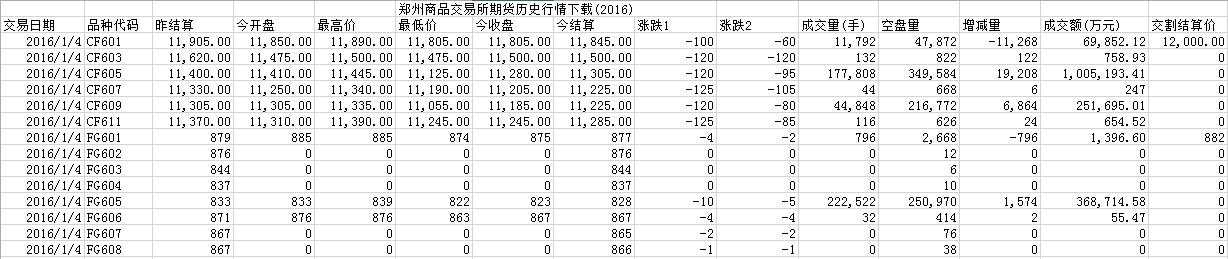


图7 郑州商品交易所2016年原始数据

由于在原始行情数据中参杂多个品种，同时每个品种每日还有多个合约，所以本文通过用Excel对数据进行筛选，选取棉花的主力合约作为实验数据，即每日成交量最大、持仓量最多、最活跃的棉花期货合约。



图8 2016年棉花期货价格总表

棉花主力合约CF605、CF705的时间跨度为2016年1月4日到2016年12月30日，共244个数据。此时每个数据中仍包含多个变量，但对投资者来说，最有参考价值、最感兴趣的是今结算，同时其他变量也与今结算有一定联系。所以本文直接选取今结算进行预测。

为了多方位地预测期货价格，本文将用现货价格预测期货价格，看这两者间的相互关系。现货价格数据来源于生意社中棉花现期图，时间跨度与期货价格数据相同，经过Excel整合后如下图。

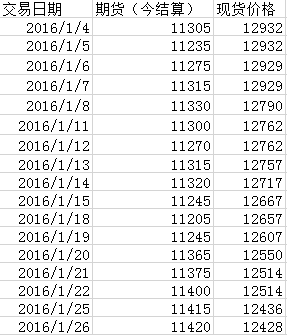


图9 棉花现货价格和期货价格汇总

为了进行预测，将棉花期货和现货数据选取前222个数据作为训练集来搭建模型，后22个数据来检验模型的预测能力。

4.2 单一模型进行棉花期货价格预测

本文的数据实证阶段采用的模型按照第二章中所列举的模型进行预测，所使用到的预测软件主要有MATLAB 2012a、Eviews 6.0等相关软件。

4.2.1 ARIMA模型进行棉花价格预测

本文选择Eviews 6.0对棉花期货结算价进行时间序列分析，在构建ARIMA模型前，要先对数据进行整体趋势的把握，由图10可以发现，棉花期货的价格呈现整体上升趋势，但变动的随机性很强，没有太大的规律性。而且数据波动幅度和变化区间都比较大。为了降低数据的波动性，使数据更加平稳，首先对其进行取自然对数，并将取对数后的结果作为原始数据，如图11，此时数据的波动幅度有所降低，变化区间相应降低。

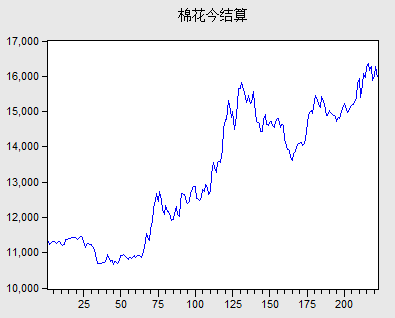


图10 棉花期货价格折线图

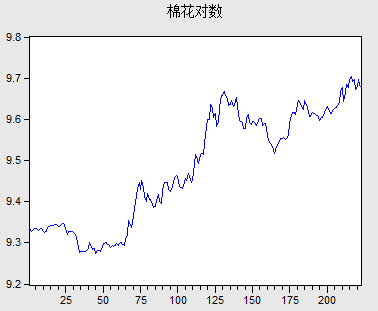


图11 棉花期货对数折线图

从图11可以看出，此时的原始数据是非平稳，为了进一步确定，对原始数据进行ADF检验，如图12所示，t值为-0.44大于在显著性水平为1%、5%、10%情况下临界值，都拒绝原假设，换而言之，原始数据并非是平稳序列。

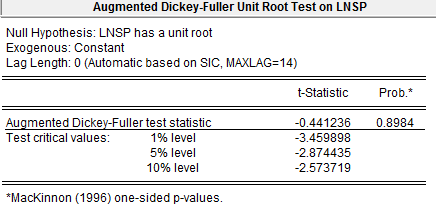


图12 棉花期货对数ADF检验

由于ARIMA模型建模时要求数据平稳，所以对数据进行一阶差分处理，得到如下图，此时可以发现数据在一个常数周围上下波动。

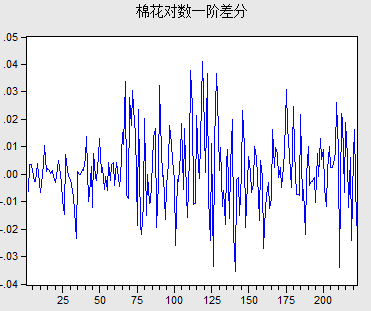


图13 棉花期货一阶差分

虽然从线形图中可以看出，进行一阶差分后的数据已经趋于平稳，但为了避免人为判断时的主观性，再次用ADF检验，如图14，此时t值为-12.74小于1%的显著性水平下的临界值，证明一阶差分后数据趋于平稳，可见d值应该取1。

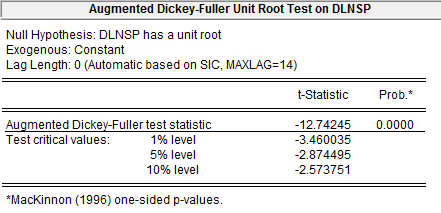


图14 棉花期货一阶差分ADF检验

正如本文在第二章所提到的，在建立ARIMA(p,d,q)模型时最重要的是确定p值和q值，同时在第二章中也提到如何更好地选择参数，所以在这个实例中，将会使用前面介绍的方法进行实验。为了更好地确定最优的p值和q值，首先绘制出一阶差分后的数据的自相关图和偏自相关图，如图15。

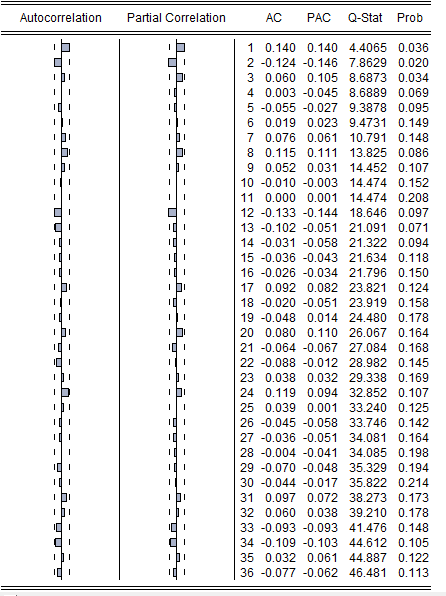


图15 棉花期货自相关图和偏自相关图

再根据观察法，可知p值应取2，q值应取2。此时模型确定为ARIMA(2,1,2)。对这个模型进行相关参数估计，如图16。

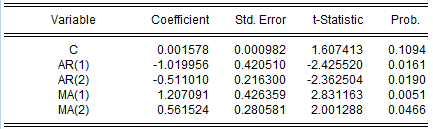


图16 ARIMA(2,1,2)参数估计图

由图可以看出仍有个别的估计值不显著，此时根据上文的方法列出可能的组合，整理后如下表：

表8 ARIMA(2,1,2)组合汇总

|  |
| --- |
| ARIMA(2,1,2)可能组合 |
| {AR(1)}; {AR(2)}; {MA(1)}; {MA(2)}; {AR(2),MA(1)};  {AR(2),MA(2)}; {AR(1),MA(1)}; {AR(1),MA(2)};  {AR(1),AR(2),MA(1)}; {AR(1),AR(2),MA(2)};  {AR(1),MA(1),MA(2)}; {AR(2),MA(1),MA(2)};  {AR(1),AR(2),MA(1),MA(2)} |

对所有列出的组合都进行参数估计，对于显著性检验不满足条件的组合进行排除，最后选取出显著性符合条件的所有组合如图17-图20：

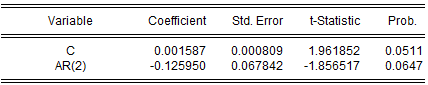


图17 AR(2)相关参数估计图

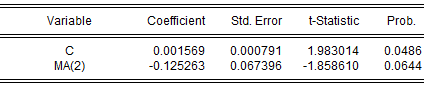


图18 MA(2)相关参数估计图

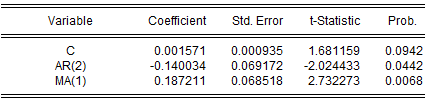


图19 AR(2)MA(1)相关参数估计图

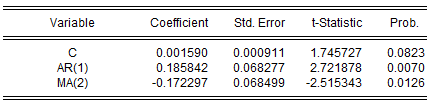


图20 AR(1)MA(2)相关参数估计图

由于有多个组合，所以通过选择各个组合中AIC值和SC值最小的作为最优组合，如果存在两个，则通过RMSE值来再次判断。四个组合的AIC和SC值如下。

表9 各组合AIC值和SC值汇总表

|  |  |  |
| --- | --- | --- |
|  | AIC值 | SC值 |
| AR(2) | -5.766724 | -5.735774 |
| MA(2) | -5.774374 | -5.743622 |
| AR(2)MA(1) | -5.790291 | -5.743865 |
| AR(1)MA(2) | -5.794570 | -5.748294 |

根据AIC和SC值的最小值的组合来确定模型，此时AR(1)MA(2)为最优。

确定模型后要对模型的残差进行白噪声检验，以确保该模型能够作为棉花期货预测的模型，画出该模型的残差自相关图和偏自相关图，如图21。可以发现该模型的残差的自相关图和偏自相关图中的柱形图都落在虚线内，说明该数据满足白噪声序列的要求，可以进行预测。

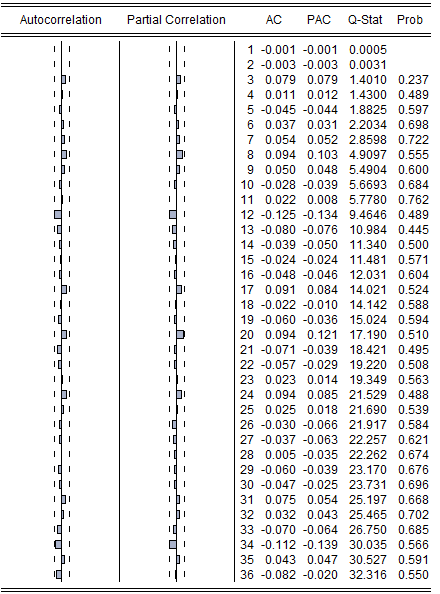


图21 棉花期货残差自相关图和偏自相关图

通过白噪声检验后，可以用该模型进行拟合和预测，拟合结果如图22。

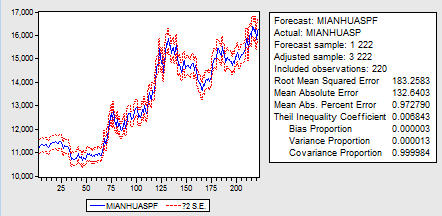


图22 棉花期货ARIMA模型拟合效果图

由图中的拟合效果图可以看出这个模型较好地模拟了棉花期货的实际变动趋势，而且从拟合的数值上看，也是拟合效果相对较好的。如Theil不等系数为0.0068说明实际值和拟合值之间的差距相对较小，协方差比为0.9998说明拟合效果也比较好。将该模型的222个合值和预测得到的22个值保存下来，作为后续分析的数据。

4.2.2 VAR模型进行棉花价格预测

本文使用Eviews 6.0进行VAR模型预测。进行VAR模型预测前，绘制棉花的现货价格和期货价格的线形图，如图23。棉花期货和现货的价格呈现整体上升趋势，变化趋势大致相同。

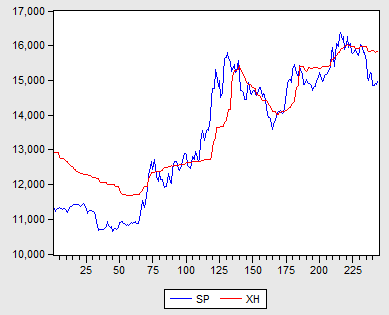


图23 棉花现货价格和期货价格折线图

由于两者的数据波动幅度和变化区间都比较大，为了后续方便处理，先对这两者进行取自然对数，减小变动区间，使数据变得平稳。并将取对数后的结果作为原始数据，如图24。

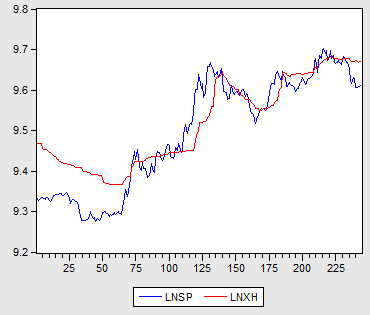


图24 棉花现货价格和期货价格对数折线图

由图24可以看出数据不平稳，为了等到更加准确的结果，先对数据进行ADF检验。

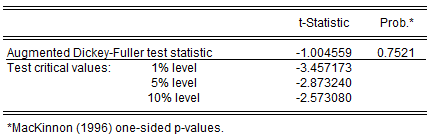


图25 棉花现货价格ADF检验

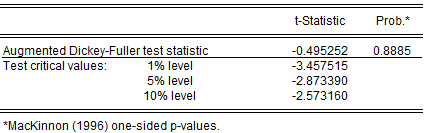


图26 棉花期货价格ADF检验

由图25和图26的ADF检验的结果可知，两者的t值都大于在显著性水平为1%、5%、10%情况下临界值，都拒绝原假设，所以棉花期货价格和现货价格是不平稳序列。

对棉花期货价格和现货价格分别进行一阶差分，并进行ADF检验，此时可以发现两者的t值都是小于显著水平为1%的临界值，此时接受原假设，即此时两者的数据已经平稳。

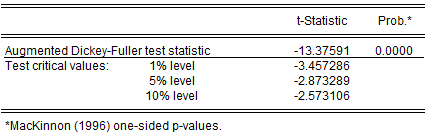


图27 棉花现货价格一阶差分后ADF检验

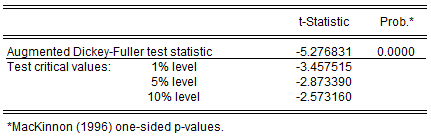


图28 棉花期货价格一阶差分后ADF检验

在数据平稳后，要确定滞后期数p，Eviews 6.0提供了多种的确认方式，同时还有自动计算滞后期数的多种方法。这里使用软件中自动选择的功能，即通过由软件使用多种计算滞后期数的方法来确定最优的滞后期数。

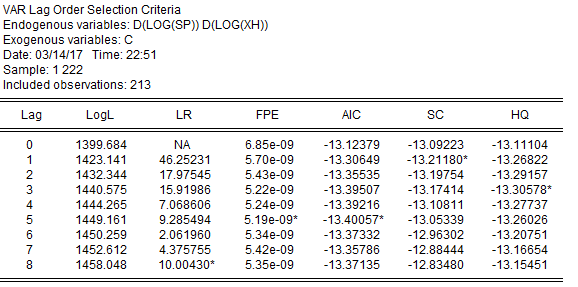


图29 Eviews自动计算最优滞后期数图

由图中可以发现软件中的有两种方法确定的滞后期数是5，其他的都是一个。所以确定p为5。

为了确认棉花的现货价格是否会对棉花的期货价格产生影响，所以在进行后续预测前，要先对数据进行格兰杰因果关系检验，来确定棉花的现货价格是否能够作为预测棉花期货价格的原因，如图30。

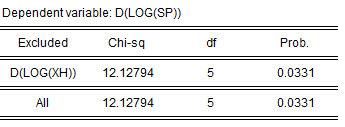


图30 棉花现货期货价格格兰杰因果关系检验

此时发现p值为0.03小于0.05，拒绝原假设，即现货价格是期货价格的格兰杰原因，因此，可以用棉花的现货价格来预测其期货价格。再通过AR ROOTS来验证模型的稳定性，确认是否可以进行VAR建模，结果如图31。

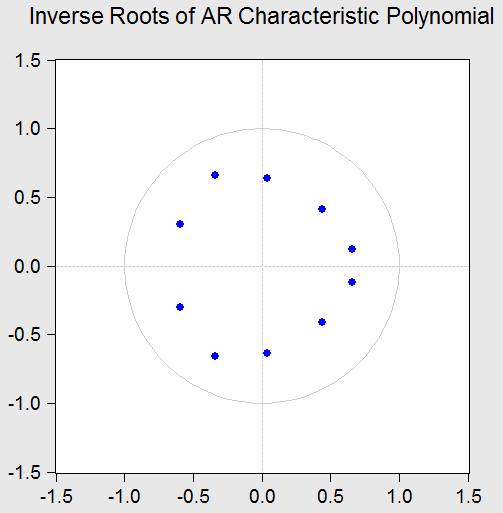


图31 AR ROOTS稳定性检验

此时发现所有的根都落在单位圆内，所以此模型是稳定的，可以进行VAR模型预测。

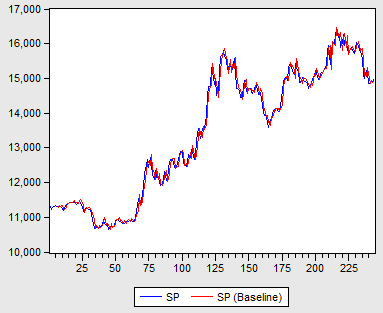


图32 棉花期货VAR模型拟合效果图

由图32可以看出这个模型较好地模拟了棉花期货的实际变动趋势，将该模型的222个拟合值和预测得到的22个值保存下来，作为后续分析的数据。

4.2.3 BP神经网络进行棉花价格预测

本文对于BP神经网络模型的搭建上，采用Matlab 2012a。BP神经网络在进行时间序列预测上最重要的是要确定其滞后的期数，由于在进行VAR建模的过程中，已经通过多种检验方式确定了滞后期数为5期，所以在BP神经网络的建模中，就默认使用5期滞后。除了滞后期数之外，神经网络的结构也是影响模型预测效果好坏的重要一环，但从第二章的BP神经网络介绍中可以知道，三层结构只要在神经元节点数和权重的分配上设计得好的情况下，也是可以很好地进行预测，所以本文在该实验中采用三层的BP神经网络。——激活函数作用、选择？，如logsig

在输入层神经元节点设计上，由于是采用5期滞后，所以在数据分配上，将前5期的棉花期货价格数据作为神经网络的输入数值，而第6期作为神经网络训练的目标数据，并按照这种方式进行滚动排列。此部分代码如下。

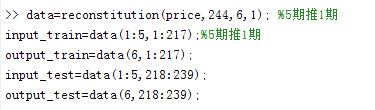


图33 神经网络输入输出设计代码

在输出层的设计上，由于本实例只有棉花期货价格一个变量，所以在输出层的神经元个数设计上为1。同时，由于期货价格数据数值较大，为了避免因为变量数量级差别太大而导致的误差，所以先将数据进行归一化处理。由于采用的是logsig作为传递函数，归一化后的区间保持在[0,1]范围内。

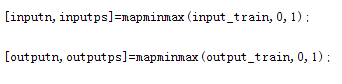


图34 BP神经网络数据归一化

设置好神经网络训练时的各类参数，如图35。

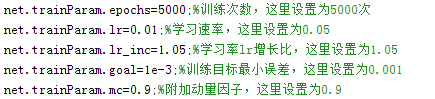


图35 BP神经网络参数设置

同时，由于神经元个数需要反复进行调试，所以在代码实现过程中，\\直接将神经元个数采用for循环的方式进行，但在区间选择上根据第三章提到的公式，得到的区间为2到12。在参数设置完毕后，开始神经网络训练，最终得到隐藏层神经元个数从2到12的误差如图36。

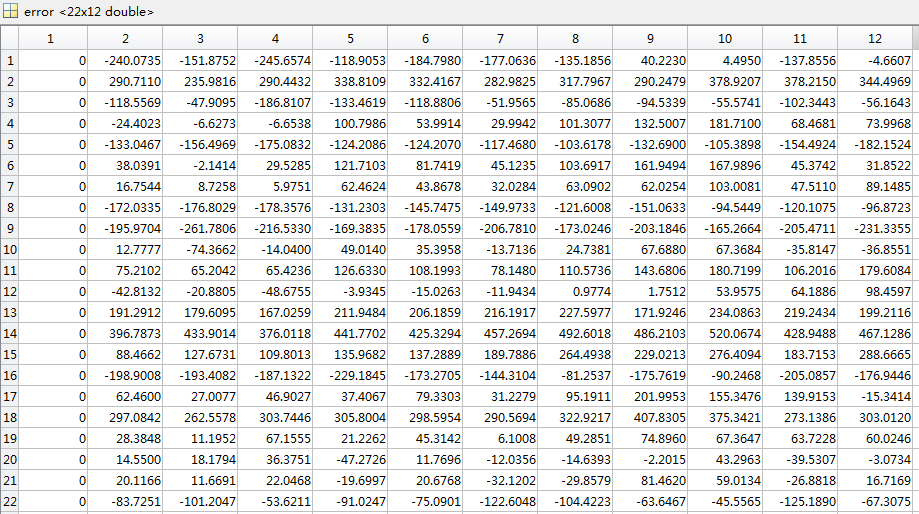


图36 2到12个神经元训练结果误差

此时难以直观地观察出最佳的神经元节点数，所以用误差绝对值百分比之和的方式来列出各个节点数的最终结果。

表10 不同神经元个数训练结果误差绝对值之和

|  |  |  |  |
| --- | --- | --- | --- |
| 神经元个数 | 误差绝对值之和 | 神经元个数 | 误差绝对值之和 |
| 2 | 17.84% | 8 | 20.64% |
| 3 | 1824% | 9 | 20.44% |
| 4 | 17.69% | 10 | 21.33% |
| 5 | 17.92% | 11 | 17.78% |
| 6 | 18.58% | 12 | 20.24% |
| 7 | 19.25% |  |  |

此时可以发现当隐藏层的神经元个数为4的时候，误差绝对值之和最小。此时表示该模型的拟合效果最好。

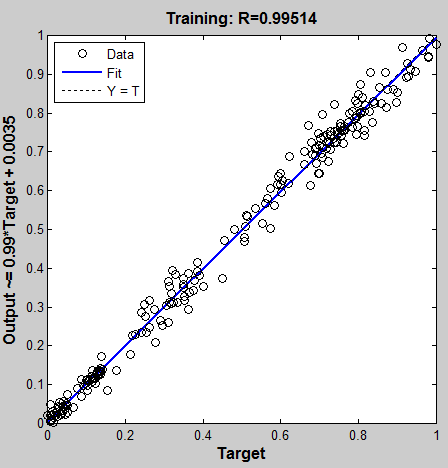


图37 隐藏层为4个神经元的神经网络拟合效果

结合训练效果的拟合图可以看出，绝大部分的值都在实际值周围，同时拟合度达到了0.99514，说明这个模型的拟合效果很好，此时将该模型的222个拟合值和22个预测值保存下来，为接下来的预测分析提供准备。

4.3 集成模型进行棉花期货价格预测

在进行集成预测前，先整合上一节三个预测模型的最终预测值和误差绝对值以方便为本节集成提供数据支持，整合后如图38。

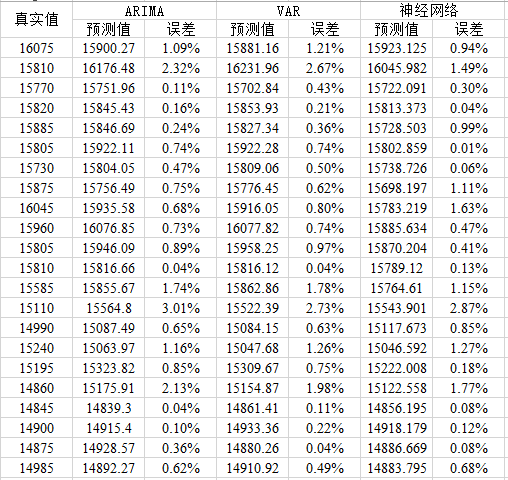


图38 单一预测方法预测值和误差

4.3.1 线性集成进行棉花价格预测

在集成前，先将三个模型在前222个数据中的拟合值的误差绝对值和误差平方和进行汇总，方便后续集成的开展。整理后如表11。

表11 单一预测方法的误差绝对值和误差平方和

|  |  |  |  |
| --- | --- | --- | --- |
|  | ARIMA | VAR | 神经网络 |
| 平均误差绝对值 | 134.6332 | 131.075 | 133.8634 |
| 误差平方和 | 7384713.5 | 6900264 | 7030660 |

对于等权集成，由于不需要分析各单一预测模型效果，所以直接根据公式可知，最终的集成表达式为：

(19)

其中，是ARIMA模型在第t期的预测结果；是VAR模型在第t期的预测结果；是BP神经网络模型在第t期的预测结果。

对于基于误差绝对值的变权预测，则要根据第三章的公式列出各个模型的权重公式，ARIMA模型的权重为=0.3290，类似地，VAR模型为0.3210，神经网络为0.3500。此时根据误差绝对值的列出的线性集成预测的表达式为：

(20)

其中，是ARIMA模型在第t期的预测结果；是VAR模型在第t期的预测结果；是BP神经网络模型在第t期的预测结果。

对于基于误差平方和的变权预测，则要根据第三章的公式列出各个模型的权重公式，ARIMA模型的权重为=0.3178，类似地，VAR模型为0.3117，神经网络为0.3705。此时根据误差绝对值的列出的线性集成预测的表达式为：

(21)

其中，是ARIMA模型在第t期的预测结果；是VAR模型在第t期的预测结果；是BP神经网络模型在第t期的预测结果。

此时根据这三种线性集成方法算出的预测结果和误差如图39。

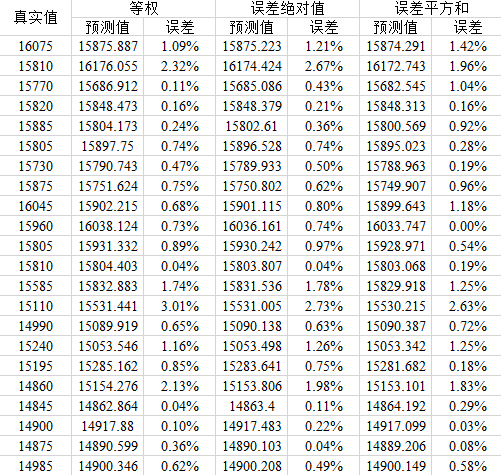


图39 线性集成预测值和误差

4.3.2 非线性集成进行棉花价格预测

在进行基于BP神经网络的非线性集成时，其实大致的操作方法与在上一小节中的神经网络预测时间序列类似，它们之间唯一的区别便是进行时间序列预测时输入值是一开始设定好的滞后期数，而在非线性集成中的输入值是多个单一预测模型的拟合值、预测值。在本实例中，训练集是上面三个单一预测模型的拟合值，预测集是三个单一预测模型的预测值。为了方便操作，首先将三个模型的拟合值和预测值进行整合。

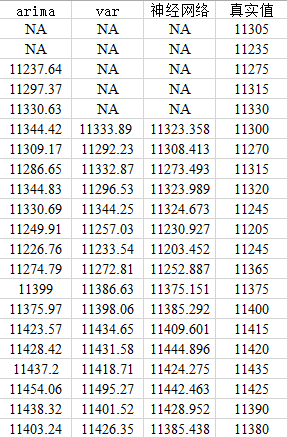


图40 单一预测方法预测值汇总

可以发现三个模型在前几期都有空值，这是因为ARIMA模型在操作时设置了两期滞后，VAR和神经网络设置了5期滞后，所以调整训练集的个数为217个，预测集为22个。同时，由于在BP神经网络的集成上依旧是选择三层结构，所以隐藏层中神经元的个数是需要根据第三章的公式进行调整，此时最佳的区间范围为3到13个神经元。对这些神经元个数进行实验得到下图的误差值。

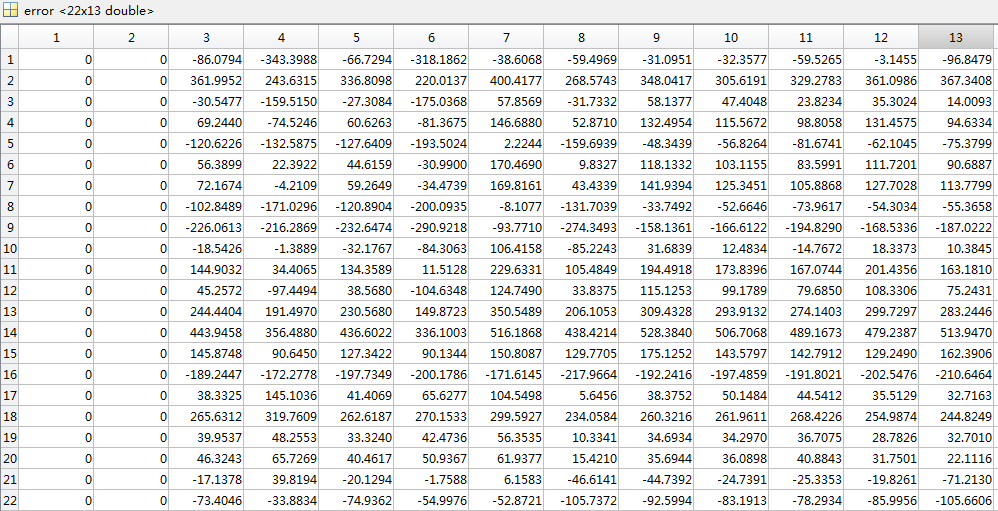


图41 3到13个神经元的神经网络集成预测误差

为了方便确定最佳神经元个数，列出每个模型的误差绝对值之和。

表12 不同神经元个数预测误差绝对值之和

|  |  |  |  |
| --- | --- | --- | --- |
| 神经元个数 | 误差绝对值之和 | 神经元个数 | 误差绝对值之和 |
| 3 | 18.04% | 9 | 17.71% |
| 4 | 18.35% | 10 | 19.01% |
| 5 | 17.74% | 11 | 18.15% |
| 6 | 17.45% | 12 | 18.53% |
| 7 | 19.21 % | 13 | 21.76% |
| 8 | 19.31% |  |  |

此时可以发现当隐藏层的神经元个数为6的时候，误差绝对值之和最小。此时表示该模型的拟合效果最好。结合训练效果的拟合图可以看出，拟合度达到了0.99511，说明这个模型的拟合效果很好。

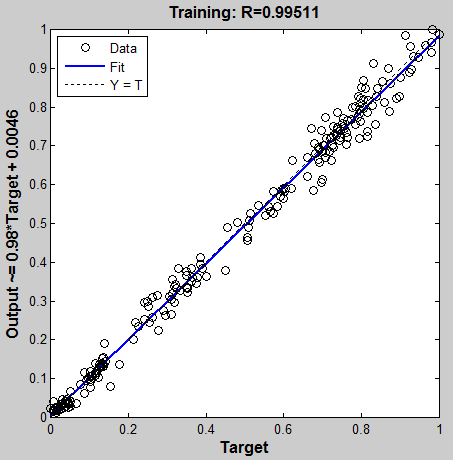


图42 隐藏层为6个神经元的神经网络集成拟合效果

本节在使用支持向量机进行集成分布将采用台湾大学林智仁教授开发设计的LIBSVM模型，该模型将支持向量机内部复杂的寻参等过程进行封装，减少使用者所需的对参数进行调节所需花费的时间。在参数选择调优中给出了三种寻参的方式，分别是网格搜索（GridSearch）、遗传算法（Genetic Algorithm，AG）和粒子群算法（Particle Swarm Optimization，PSO），使用者可以通过选择不同的寻参方式来寻找最适合所应用数据的参数，同时在最优参数的选择上，该模型通过对得到的参数组合进行交叉检验来选取出预测准确度最高的参数组合作为最终的输出结果。

在本次实验中，分别使用这三种寻参的方式进行实验，通过选取出结果误差绝对值最小的结果作为支持向量机集成结果。

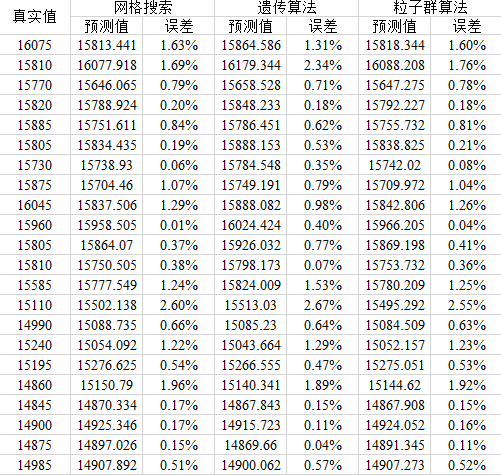


图43 三种支持向量机集成预测值和误差

从集成结果看，这几个模型都存在一定的波动，同时在其中特定的值中误差波动会比较大，这也与原有的单一预测方法在预测时所呈现的结果有一定的关系。这里采用误差绝对值之和的方式来确定出最佳的寻参算法。

表13 三种支持向量机误差绝对值之和

|  |  |  |  |
| --- | --- | --- | --- |
| 寻参算法 | 网格搜索 | 遗传算法 | 粒子群算法 |
| 误差绝对值之和 | 17.72% | 18.38% | 17.57% |

从结果中可以看出，粒子群算法的误差绝对值之和最小，即选取粒子群算法作为支持向量机的寻参算法。

4.3.3 实例预测结果分析与结论

经过的单一预测方法和集成预测方法之后，各种模型均得到了22个预测结果值，预测结果和预测误差绝对值汇总后如图44。



图44 单一预测方法和集成预测方法预测结果汇总

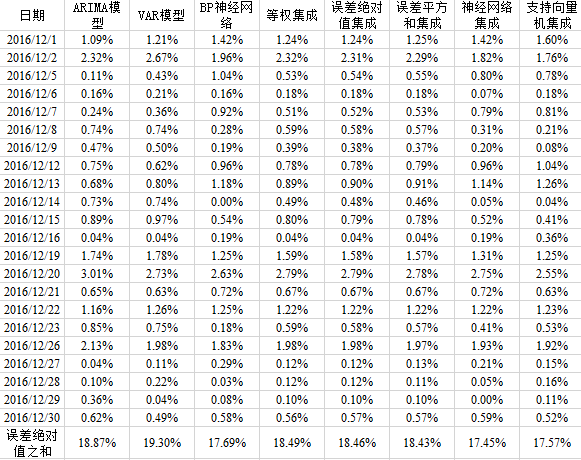


图45 单一预测方法和集成预测方法预测误差汇总

由图45的最后一行可以初步看出所有的结果中，单一预测模型中神经网络的预测效果最好，这与神经网络的结构选择较优有一定的关系。在集成方面，线性集成能够较好地整合三个单一方法的预测结果，虽然没有达到神经网络的预测效果，但是整体上融合三个方法的优点。而非线性集成的预测效果优于线性集成的效果，且都由于神经网络的集成效果。

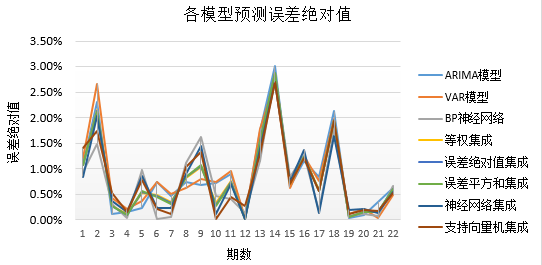


图46 单一预测方法和集成预测方法误差折线图

由图46中可以大致看出，各个模型在预测过程中总体的误差波动趋于一致，这表明这些模型在预测整体趋势上表现较为相近，但也能发现在一些期数上的波动较大，这可能因为在这些时间点中棉花的期货市场受到了其他外部条件的影响。

将所有模型的残差平方和、平均绝对误差和平均绝对百分比汇总成表，进行进一步分析。

表14 单一预测方法和集成预测方法结果评估汇总表

|  |  |  |  |
| --- | --- | --- | --- |
|  | 残差平方和 | 平均绝对误差 | 平均绝对百分比误差 |
| ARIMA模型 | 694826.24 | 132.63 | 0.858% |
| VAR模型 | 708478.54 | 135.95 | 0.877% |
| BP神经网络 | 595967.29 | 124.68 | 0.804% |
| 等权集成 | 644469.41 | 130.15 | 0.840% |
| 误差绝对值集成 | 642218.05 | 129.97 | 0.839% |
| 误差平方和集成 | 639626.90 | 129.77 | 0.838% |
| 神经网络集成 | 589834.43 | 122.86 | 0.793% |
| 支持向量机集成 | 578148.20 | 123.77 | 0.798% |

从所有的数值上看，对于ARIMA模型、VAR模型、BP神经网络这三个模型在这三种评价标准上的都是BP神经网络的结果最优，其次是ARIMA模型，最后是VAR模型。在线性集成方面，误差平方和集成的各项数值均最小，表明其集成效果最好，其次是误差绝对值集成，而等权集成由于没有考虑到各个模型的预测优劣，所以效果没有那么突出。但整体上线性集成预测仍是比单一预测的效果更加稳定。在非线性集成方面，神经网络集成在残差平方和的表现上没有支持向量机表现突出，在平均误差值和平均绝对百分比误差上略优于支持向量机，主要是因为神经网络在集成过程中一些数值误差较大，但整体误差更小，所以在神经网络集成整体效果最好，但是支持向量机在预测稳定性更强。

综上，使用集成预测，特别是非线性集成预测在预测上的表现会优于单一预测方法，可以将本文的研究成果运用到其他不同的领域，可以起到一定的参考意义。

5 总结与展望

5.1 总结

预测在日常生活中的方方面面起着越来越大的作用，对预测方法的研究是提高预测结果精确度的一个重要的方法，而集成预测是为了更好地融合各个模型在预测上的优点，提高最终预测结果的准确度而提出的，所以对集成预测方法进行研究就显得非常必要和有意义。

本文首先通过对现有的单一预测方法进行总体的介绍后，再列举出比较有代表性的三种预测方法（ARIMA模型、VAR模型、BP神经网络模型）进行介绍，其中在ARIMA模型的介绍中还涉及一种新的、更加科学的p值和q值的确定方法，这让预测精度得到一定的提升。在后面的集成预测部分，本文则通过对集成预测方法的不同进行分类，分为线性集成和非线性集成。在线性集成中以最常规的等权、基于误差绝对值、基于误差平方和这三种进行介绍展开，在非线性集成中以基于神经网络和基于支持向量机进行展开，并通过一个实例进行简单地实例验证，证明出集成预测的结果会优于单一预测的结果。

在介绍完理论部分后，本文又采用备受关注的农产品期货中的棉花期货数据作为实验的预测对象，通过使用理论部分谈及的ARIMA模型、VAR模型和BP神经网络分别进行了预测，其中神经网络的预测效果最优。再将这三种模型的按照理论部分分别进行线性集成和非线性集成，得到的预测结果表明在农产品期货中，集成预测的结果能够起到很好的预测指导作用，同时也能够很好地综合各个单一预测模型的预测效果，所以这也表明集成预测在集成效果上有着优秀的表现。以后在农产品期货价格预测中进行集成预测也是一种很好的预测途径。

5.2 展望

本文虽从线性与非线性的角度对集成预测进行了相关的分析，但由于本人在知识储备、研究能力和实践能力上还有所欠缺，在集成部分的讲述和实践中未能够采用到国际先进的集成预测方法，如小波分析、EMD等，同时在集成方式上也局限在只是对结果进行集成，没有尝试在整个预测过程中进行集成，这是不足之处。

此外在进行农产品期货预测上由于时间关系，只是对算法进行了解使用，没有尝试进行优化，以及信息不对称的原因，也不能够找到更多纬度的数据进行分析，造成预测结果上仍存在一些较大的误差，这是比较遗憾的地方。

以上的这些问题也是需要在今后进行改进和完善的，相信随着时代的不断发展，有关学者在集成预测方法上的不断深入，集成预测一定会更受重视，其预测精度也会越来越高。

参 考 文 献

陈华友.组合预测方法有效性理论及其应用[M].北京：科学出版社，2008：20-25.

孔庆凯.平均预测法的应用条件[J].预测，1985（5）：18-21.

唐小我.组合预测计算方法研究[J].预测，1991（4）：35-39.

唐小我，曾勇，曹长修.变权组合预测摸型研究[J].预测，1993（3）：46-48.

邓雪，唐焕文，陈桂枝.组合预测误差平方和取值范围研究[J].大连理工大学学报，2001（6）：637-639.

高尚，张绍彪，梅亮.基于相对误差的线性组合预测研究[J].系统工程与电子技术，2008，3（25）：481-484.

文新辉，牛明洁.一种新的基于神经网络的非线性组合预测方法[J].系统工程理论与实践，1994（12）：66-72.

董景荣.基于小波网络的非线性组合预测方法研究[J].系统工程学报，2000（12）：383-388.

李元诚，李波，方廷健.基于小波支持向量机的非线性组合预测方法研究[J].信息与控制，2004，33（3）：303-306.

耿建军，吕卫民，董景荣.基于GRNN神经网络的变权组合预测的权重确定方法[J].数学的实践与认识，2011，41（3）：86-93.

芮兰兰，李钦铭.基于组合模型的短时交通流量预测算法[J].电子与信息学报，2016，38（5）：1227-1233.

王江，费宇.上海锌期货价格的组合预测分析[J].中国证券期货，2011（1）：10-12.

褚东.时间序列模型对大豆期货价格的预测比较研究[J].金融经济，2012：116-119.

陈林，黄章树.基于ARIMA模型的期货价格分析与预测[J].福州大学学报，2010，24（3）：32-37.

王凤飞.组合预测方法简介及其实证分析[D].山西大学，2011.

J.M.Bates，C.W.J.Granger.The Combination of Forecasts[J].Journal of the Operational Research Society，1969，20（4）：451-468.

S Makridakis，RL Winkler.Averages of Forecasts: Some Empirical Results[J].Management Science，1983，29（9）：987-996.

Diebold F X.Serial correlation and the combination of forecasts[J].Journal of Business and Economic Statisties，1988，6：105-111.

Crawford G W，Fratantoni M C.Assessing the Forecasting Performance of Regime-Switching，ARIMA and GARCH Models of House Prices[J]. Real Estate Economics，2003，31（2）：223-243.

David E.Rapach，Jack K.Strauss.Differenees in housing price Forecastability across US States [J].International Journal of Foreeasting，2009，2（25）：351-372.

Granger C W J.Combining Forecasts-Twenty Years Later[J].Journal of Forecasting，1989，8 （3）：167-173.

DG Anandalingam，DL Chen.Linear combination of Forecasts:A General Bayesian Model[J].Journal of Forecasting，1989，8（3）：19-214.

Guerard J B，Clemen R T.Collinearity and the Use of Latent Root Regression for Combining GNP Forecasts[J].Journal of Forecasting，1989，8（3）：231-238.

附 录

**附录1 BP神经网络代码**

data=reconstitution(price,244,6,1); %5期推1期

input\_train=data(1:5,1:217);%5期推1期

output\_train=data(6,1:217);

input\_test=data(1:5,218:239);

output\_test=data(6,218:239);

[inputn,inputps]=mapminmax(input\_train,0,1);

[outputn,outputps]=mapminmax(output\_train,0,1);

for i=2:12 %从2到12个神经元开始遍历

net=newff(minmax(inputn),[i,1],{'logsig','purelin'},'trainlm');

net.trainParam.epochs=5000;%训练次数，这里设置为5000次

net.trainParam.lr=0.01;%学习速率，这里设置为0.05

net.trainParam.lr\_inc=1.05;%学习率lr增长比，这里设置为1.05

net.trainParam.goal=1e-3;%训练目标最小误差，这里设置为0.001

net.trainParam.mc=0.9;%附加动量因子，这里设置为0.9

[net,tr]=train(net,inputn,outputn);

inputn\_test=mapminmax('apply',input\_test,inputps);

an=sim(net,inputn\_test);

BPoutput=mapminmax('reverse',an,outputps);

%得到训练结果

an\_train=sim(net,inputn);

%训练结果反归一

BPoutput\_test=mapminmax('reverse',an\_train,outputps);

BPoutput\_train(:,i)=BPoutput\_test';

figure(i)

plot(BPoutput,':og')

hold on

plot(output\_test,'-\*');

legend('预测输出','期望输出')

title('BP网络预测输出','fontsize',12)

ylabel('函数输出','fontsize',12)

xlabel('样本','fontsize',12)

%预测误差

predict(:,i)=BPoutput';

error(:,i)=BPoutput'-output\_test';

end

**附录2 支持向量机代码**

tr\_x\_scale=data(1:3,1:216)';

tr\_y\_scale=data(4,1:216)';

testx=data(1:3,217:238)';

testy=data(4,217:238)';

Method\_option.plotOriginal = 0;

Method\_option.xscale = 1;

Method\_option.yscale = 1 ;

Method\_option.plotScale = 0;

Method\_option.pca = 0 ;

Method\_option.type = 1;

[predict\_Y,mse,r] = SVR(tr\_y\_scale,tr\_x\_scale,testy,testx,Method\_option);

predict=predict\_Y{:,2};

error(:,i)=real\_high-predict;

end

致 谢

在进行毕业论文编写的这段时间是我大学中最有意义、最有价值的时光。借着论文写作完成之际，我想要对所有帮助过我的人表示我的感谢。

这段时间中，我最想要感谢的是我的指导老师——张大斌老师，张老师在我的整个毕业论文的写作过程中，无论是在课题选择、研究思路、研究方式方法上都给了我非常多的指导，正因为与他多次的沟通交流才让我的毕业课题有了更加丰富的内容。

同时，我也要感谢在这段时间中同班的同学的帮助，正是有他们与我在思想上不断的交流碰撞才能够让我在编写的过程中高效地完成我的论文写作。

再者，我还要感谢在大学四年间我的授课老师们，正是有他们在学习上的教导帮助，才让我学到各方面的专业知识和在自主学习的习惯。他们身上严格的治学工作态度也深深地感染了我。正因为他们四年来对我潜移默化的影响，才让我能顺利完成我的论文。

最后，再一次向我的老师、同学以及我的家人朋友们致以最真挚的感谢。

**附件7**

**华南农业大学**

**本科生毕业论文成绩评定表**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学号 |  | | | | | 姓名 | |  | 专业 |  | | |
| 毕业论文题目 | | |  | | | | | | | | | |
| 指导教师评语  成绩（百分制）：　　　　　　　　　　　　　　　　　　　　指导教师签名：　　　　　　　　　　　　　　　年　　月　　日 | | | | | | | | | | | | |
| 评  阅  人  评  语  及  成  绩  评  定 | | 成绩  评定  标准 | | **评分项目** | | | | | | | **分值** | **得分** |
| 选题  质量  20% | 1 | | 专业培养目标 | | | | 5 |  |
| 2 | | 课题难易度与工作量 | | | | 10 |  |
| 3 | | 理论意义或生产实践意义 | | | | 5 |  |
| 能力  水平40% | 4 | | 查阅文献资料与综合运用知识能力 | | | | 10 |  |
| 5 | | 研究方案的设计能力 | | | | 10 |  |
| 6 | | 研究方法和手段的运用能力 | | | | 10 |  |
| 7 | | 外文应用能力 | | | | 10 |  |
| 成果  质量40% | 8 | | 写作水平与写作规范 | | | | 20 |  |
| 9 | | 研究结果的理论或实际应用价值 | | | | 20 |  |
| 评阅人评语：  成绩（百分制）：　　　　　　　　　　　　　评阅人签名：　　　　　　　　　　　　年　　月　　日 | | | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 答  辩委员会  意见与  成  绩  评  定 | 评价项目 | 具体要求（A级标准） | 最高分 | 评分 | | | | |
| A | B | C | D | E |
| 论文质量 | 论文（设计）结构严谨，逻辑性强；有一定的学术价值或实用价值；文字表达准确流畅；论文格式规范；图表（或图纸）规范、符合要求。 | 60 | 55-60 | 49-54 | 43-48 | 37-42 | ≤36 |
|  |  |  |  |  |
| 论文报告、讲解 | 思路清晰；概念清楚，重点（创新点）突出；语言表达准确；报告时间、节奏掌握好。 | 20 | 19-20 | 17-18 | 15-16 | 13-14 | ≤12 |
|  |  |  |  |  |
| 答辩情况 | 答辩态度认真，能准确回答问题 | 20 | 19-20 | 17-18 | 15-16 | 13-14 | ≤12 |
|  |  |  |  |  |
| 是否同意通过论文答辩（打√）   1. 同意 2. 不同意   成绩（百分制）：　　　　　　　　 答辩委员会主席（签名）：  　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　 年 月 日 | | | | | | | |

续上表：

|  |  |
| --- | --- |
| 成绩总评 | 论文总评分数：  论文成绩总评等级：　　　 学院盖章：  年 月 日 |

**注：1、**论文成绩评定等级：参考评阅人的评阅、指导教师评阅情况，结合答辩情况，建议按指导教师评分、评阅人评分、答辩评分为4∶3∶3的比例评定论文总成绩分数，然后按优（90－100分）、良（80－89分）、中（70－79分）、及格（60－69分）、不及格（<60分）给出成绩等级。**2**、论文成绩以分数的形式登记到教务管理系统。

**附件8 华南农业大学**

**本科生毕业设计成绩评定表**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 学号 | |  | | 姓名 |  | 专业 |  |
| 毕业设计题目 | | |  | | | | |
| 指导教师评语  成绩（百分制）：　　　　　　　　　　　　　　　指导教师签名：　　　　　　　　　　　年　　月　　日 | | | | | | | |
| 评阅人评语  成绩（百分制）：　　　　　　　　　　　　　　　评阅人签名：　　　　　　　　　　　　年　　月　　日 | | | | | | | |
| 答辩委员会评语  成绩（百分制）：　　　　　　　　　　　　　　　答辩委员会签名：　　　　　　　　　　年　　月　　日 | | | | | | | |
| 成绩总评 | 毕业设计总评分数：  毕业设计成绩总评等级：　　　 答辩委员会主席（签名）：  年 月 日 | | | | | | |

**注：1、**毕业设计成绩评定等级：参考评阅人的评阅、指导教师评阅情况，结合答辩情况，建议按指导教师评分、评阅人评分、答辩评分为4∶3∶3的比例评定论文总成绩分数，然后按优（90－100分）、良（80－89分）、中（70－79分）、及格（60－69分）、不及格（<60分）给出成绩等级。**2**、论文成绩以分数的形式登记到教务管理系统。