类号 密级

UDC

学 位 论 文

基于量子粒子群优化的电力负荷预测方法研究

|  |  |  |  |
| --- | --- | --- | --- |
| 作者姓名： | 王祝兵 | | |
| 指导教师： | 焦明海 副教授 | | |
|  | 东北大学计算中心 | | |
| 申请学位级别： | 硕士 | 学科类别： | 工学 |
| 学科专业名称： | 计算机软件与理论 | | |
| 论文提交日期： |  | 论文答辩日期： |  |
| 学位授予日期： |  | 答辩委员会主席： |  |
| 评阅人： |  | | |

东 北 大 学

[**A Thesis i**](file://C:\Users\Administrator\Desktop\论文\后面还有注解(1)Control%20Theory%20and%20Engineering(2)Discipline%20of%20ComputerSoftware%20and%20Theory%20(3)ComputerTechnology(4)ComputerOrganization%20and%20Architecture%20(5)Measuring%20and%20Testing%20Technologies%20and%20Instruments%20(6)PowerSystems%20and%20Automation)**n Computer Technology**

**Research on power load forecasting method based on quantum behaved particle swarm optimization**

By Wang Zhubing

Supervisor: Professor Jiao Minghai

**Northeastern University**

**2016.12**

# 独创性声明

本人声明，所呈交的学位论文是在导师的指导下完成的。论文中取得的研究成果除加以标注和致谢的地方外，不包含其他人己经发表或撰写过的研究成果，也不包括本人为获得其他学位而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均己在论文中作了明确的说明并表示谢意。

学位论文作者签名：

日 期：

学位论文版权使用授权书

本学位论文作者和指导教师完全了解东北大学有关保留、使用学位论文的规定：即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。本人同意东北大学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、交流。

作者和导师同意网上交流的时间为作者获得学位后：

半年 □ 一年□ 一年半□ 两年□

学位论文作者签名： 导师签名：

签字日期： 签字日期

# 摘 要

电力负荷预测一直是电厂供应侧运行、控制、调度的关键问题，电力负荷预测的准确性则会直接影响着调度结果，进而对电力系统的安全稳定生产和国民经济带来重要的影响。当前研究领域里，中长期电力负荷预测一直是研究的热点问题，从简单的时间递推的预测方法，发展为现今需要考虑多种影响因素进行预测。与此同时，随着电力大数据的到来，电力系统中迫切需要一种能够处理基于大数据的电力负荷短期预测模式。在预测模型的选取过程中特别是运用神经网络模型，它能很好的构建多种因素影响下的网络，但是会出现收敛性、准确性、鲁棒性等问题。针对上述神经网络模型预测电力负荷的缺点，本文将量子行为引入电力负荷预测中，解决中长期电力负荷预测出现的相关问题，接着提出基于并行化量子粒子群优化神经网络相关算法，对短期电力负荷进行预测。本文主要贡献如下：

1. 针对基于粒子群的神经网络算法在处理电力负荷预测方面的不足，提出改进策略，设计出基于量子搜索粒子群的神经网络改进算法，并通过与相关算法在精度，迭代次数等方面进行对比实验。
2. 设计出基于MapReduce粒子群的神经网络算法，解决传统算法在处理数据能力上的不足并对其深入研究优化，提高处理电力负荷数据效率。
3. 运用量子算法改进基于粒子群的神经网络算法，实现分布式算法，比较算法在处理电力大数据时的优势。
4. 在Hadoop平台下针对不同的分布式优化BP算法，分别对不同的数据集进行性能测试实验，并对算法运行时间进行分析。

通过对中长期电力负荷预测实验表明，本文提出的带有线性递减的收敛因子的量子粒子群优化神经网络能够在收敛速度、预测准确性上有所提高，同时在处理电力大数据的短期电力负荷预测实验表明，并行化本文提出相关算法能增强处理数据的效率。

**关键词**：电力负荷；量子算法；粒子群；神经网络 ；hadoop

# Abstract

Power load forecasting has always been a key problem in power supply side operation, control and dispatching. The accuracy of power load forecasting will directly affect the results of dispatching, which will have important influence on the safe stable production of power system and the national economy. In the current research field, medium and long-term power load forecasting has been a hot research topic, developed from the simple time-recursive forecasting method to the present situation need to consider a variety of factors to predict. At the same time, with the arrival of large power data, there is an urgent need to deal with short-term power load forecasting model based on large data in the power system. In the selection of prediction model, especially the use of neural network model, it can well build a variety of factors under the influence of the network, but there will be some problems , for example, convergence, accuracy and robustness. In order to solve the shortcomings of the neural network model, this paper introduces the quantum behavior into the power load forecasting to solve the related problems of the medium and long term load forecasting. Then, this paper put forward neural network optimization based on the paralleled quantum-patical swarm algorithm. The main contributions of this paper are as follows:

1. For neural network algorithm based on particle swarm in the process insufficient power load forecasting, this paper proposes an improved strategy, which is based on quantum search particle swarm algorithm, and compares it to the correlation algorithm on accuracy, iteration number, etc.

2. Design the particle swarm-neural network algorithm based on MapReduce to solve the lack of traditional algorithms in the data processing capacity, improve the efficiency of data processing power load.

3. Using the quantum algorithm to improve the neural network algorithm based on particle swarm, realize the distributed algorithm, compare the advantage of the algorithm in dealing with the big power data.

4. Under the Hadoop platform, different performance of different data sets was tested for different distributed optimization BP algorithms, and analyze the running time of different algorithms .

Based on the long-term power load forecast experiments show that the quantum particle decreases with linear convergence factor of the proposed swarm optimization neural network which can improve the accuracy in predicting the convergence speed, at the same time, in the short-term load forecasting of power big data processing experiments show that the parallel algorithm is proposed in this paper can enhance the efficiency of data processing.

**Keywords:** Power load forecasting ; quantum; PSO; BP; Hadoop

# 目录

[独创性声明 I](#_Toc469318244)

[摘 要 III](#_Toc469318245)

[Abstract IV](#_Toc469318246)

[目录 VI](#_Toc469318247)

[第1章 绪 论 - 1 -](#_Toc469318248)

[1.1 研究背景 - 1 -](#_Toc469318249)

[1.2 研究目的和意义 - 2 -](#_Toc469318250)

[1.3 国内外研究现状 - 3 -](#_Toc469318251)

[1.3.1 传统负荷预测 - 3 -](#_Toc469318252)

[1.3.2 新兴负荷预测 - 5 -](#_Toc469318253)

[1.4 论文章节安排 - 6 -](#_Toc469318254)

[第2章 相关技术介绍 - 8 -](#_Toc469318255)

[2.1 神经网络算法 - 8 -](#_Toc469318256)

[2.1.1 反向传播神经网络模型 - 8 -](#_Toc469318257)

[2.1.2 反向传播神经网络推导过程 - 9 -](#_Toc469318258)

[2.1.3 BP神经网络算法描述过程 - 11 -](#_Toc469318259)

[2.1.4 BP神经网络影响因素分析 - 11 -](#_Toc469318260)

[2.1.5 BP神经网络常见的优化算法 - 12 -](#_Toc469318261)

[2.2 基本粒子群算法 - 13 -](#_Toc469318262)

[2.2.1 算法起源 - 13 -](#_Toc469318263)

[2.2.2 粒子群算法原理 - 14 -](#_Toc469318264)

[2.2.3粒子群算法实现步骤 - 14 -](#_Toc469318265)

[2.2.4粒子群算法相关改进 - 15 -](#_Toc469318266)

[2.3 量子行为的粒子群算法 - 16 -](#_Toc469318267)

[2.3.1 QPSO模型建立 - 17 -](#_Toc469318268)

[2.3.2 量子位置 - 17 -](#_Toc469318269)

[2.3.3 量子粒子群的特点 - 18 -](#_Toc469318270)

[2.4 分布式平台 - 18 -](#_Toc469318271)

[2.4.1 - 18 -](#_Toc469318272)

[2.4.2 - 19 -](#_Toc469318273)

[2.4.3 编程模型 - 20 -](#_Toc469318274)

[2.5 - 21 -](#_Toc469318275)

[第3章 量子粒子群神经网络优化电力负荷预测 - 22 -](#_Toc469318276)

[3.1 神经网络在电力负荷预测中的缺点 - 22 -](#_Toc469318277)

[3.2基于PSO优化神经网络电力负荷预测 - 23 -](#_Toc469318278)

[3.2.1粒子群优化神经网络的可行性 - 23 -](#_Toc469318279)

[3.2.2粒子群神经网络关键问题 - 24 -](#_Toc469318280)

[3.2.3粒子群神经网络算法步骤 - 25 -](#_Toc469318281)

[3.3 基于量子行为的粒子群优化神经网络电力负荷预测 - 26 -](#_Toc469318282)

[3.3.1量子粒子群对神经网络优化 - 27 -](#_Toc469318283)

[3.3.2量子粒子群神经网络算法步骤 - 28 -](#_Toc469318284)

[3.4 基于改进的量子粒子群优化神经网络电力预测 - 29 -](#_Toc469318285)

[3.4.1改进的量子粒子群神经网络参数优化 - 29 -](#_Toc469318286)

[3.4.2 几种粒子群优化算法对比仿真实验 - 31 -](#_Toc469318287)

[3.5 本章小结 - 33 -](#_Toc469318288)

[第4章 分布式神经网络优化算法的电力负荷预测 - 34 -](#_Toc469318289)

[4.1 分布式神经网络优化研究的必要性分析 - 34 -](#_Toc469318290)

[4.2 分布式BP神经网络算法步骤 - 35 -](#_Toc469318291)

[4.3 分布式神经网络优化算法实现 - 37 -](#_Toc469318292)

[4.3.1 分布式BP神经网络优化算法关键问题 - 38 -](#_Toc469318293)

[4.3.2 分布式粒子群优化BP神经网络实现步骤 - 38 -](#_Toc469318294)

[4.3.3 分布式改进的量子粒子群优化神经网络算法 - 42 -](#_Toc469318295)

[4.4 本章小结 - 43 -](#_Toc469318296)

[第5章 电力负荷预测建模和实验 - 44 -](#_Toc469318297)

[5.1 数据来源及其预测内容概述 - 44 -](#_Toc469318298)

[5.2 评价标准 - 45 -](#_Toc469318299)

[5.3 电力负荷模型建立 - 46 -](#_Toc469318300)

[5.3.1 输入关联度预处理 - 46 -](#_Toc469318301)

[5.3.2 数据归一化处理 - 47 -](#_Toc469318302)

[5.3.3 参数设置 - 48 -](#_Toc469318303)

[5.4 中长期电力负荷相关优化算法实验 - 50 -](#_Toc469318304)

[5.5 短期电力负荷分布式优化算法实验 - 53 -](#_Toc469318305)

[5.5.1 分布式平台搭建 - 54 -](#_Toc469318306)

[5.5.2 模型建立与实验分析 - 55 -](#_Toc469318307)

[5.6本章小结 - 57 -](#_Toc469318308)

[第6章 总结与展望 - 59 -](#_Toc469318309)

[6.1 总结 - 59 -](#_Toc469318310)

[6.2 展望 - 60 -](#_Toc469318311)

[参考文献 - 61 -](#_Toc469318312)

[致 谢 - 64 -](#_Toc469318313)

[攻读硕士学位期间的论文和项目情况 - 65 -](#_Toc469318314)

# 第1章 绪 论

## 1.1 研究背景

电力系统负荷预测是在充分考虑现有的条件下，研究和利用一套系统的方法，在满足一定精度要求的意义下，确定某个特定时刻的负荷数值。它关系到是否能及时掌握负荷的发展趋势；是否能正确贯彻，执行国家各个时期的方针政策；是否能适应各个时期经济发展和社会进步对电力的需求；是否能保证电网的安全，经济可靠运行；是否能安排好系统的运行方式和检修计划等。

随着对电力负荷预测的需求增加，它对电力系统相关部门的重要性不言而喻。电力负荷预测包括中长期电力负荷预测和短期电力负荷预测两种类型，其中中长期电力负荷预测又分为月度电力负荷预测，年度电力负荷预测两种。年度和月度负荷预测是电力系统规划的重要工作内容，为发电厂建设提供了一种宏观的依据。短期电力负荷预测一般是对周、天、数小时的电力负荷变化情况进行合理的估计，从而能够准确制作出发电计划和定价功能。

准确的电力负荷预测是电力系统科学规划决策的基础，同时随着电网大规模的扩展和电力市场化改革深化，对电力负荷预测精度提出了更高的要求，而且随着电力数据的积累必然会形成大数据量、影响电力负荷预测因素复杂等特点，因而对其准确而高效的预测成为了电力领域的一个热点问题。长期以来，国内外的学者对电力负荷预测做了大量的研究，提出过许多预测模型，如多元线性回归、自回归滑动平均趋势外推等方法。这些方法在准确度方面不尽人意，主要因为电力负荷预测受多种因素的影响，如天气、国家的经济、大型企业的产能等。另外随着大数据的出现，目前还很少有能处理基于大数据电力负荷预测的方法。因而需要在研究过程中不仅要注重寻找到能提高预测的准确性算法而且要在大数据情况下对电力负荷预测的高效性有所突破。

由于电力负荷受到多种因素的影响，比如钢厂电力消耗、天气气候的影响、国家经济的变化等，各种因素使得负荷预测工作遇见了很多的困难。电力负荷预测的生命力就在于预测的准确性。负荷预测结果的准确性将直接影响调度的结果，从而对电力系统的安全稳定运行和经济型带来重要的影响。当前市场化运行的条件下，由于市场各方对信息的获取和运行的经济性更加的重视，准确的预测对于提高电力经营主体的运行效益有直接的作用，因而电力负荷预测的重要性也更加的突出了。另外的一种趋势越来越明显即随着电力大数据的到来，电力系统中迫切需要一种能够基于大数据处理的电力负荷的处理模式。这就需要我们解决两方面的问题：第一，电力负荷的准确性，第二，大数据下的电力预测的高效性。

## 1.2 研究目的和意义

为了优化用电结构，调整负荷曲线，保证电网安全性运行，并且充分的借助电力资源，使得成本得到降低，准确的预测电力负荷预测尤为重要。另外如何高效的对大数据趋势预测电力负荷也越来越成为热点问题。当今电力负荷预测研究中，已经取得了大量的研究成果[3,4]，很多精确度比较高的智能算法也应用到具体问题中。但这些研究大部分是在因素比较少的情况下如只考虑天气，居民电力消耗量[5]，或是没有考虑国家宏观调控因素如GDP，效果也不尽相同。另外在电力系统中随着大数据的出现，使得原来只处理小规模数据的电力系统预测起来变得越来越吃力。通常情况对大数据量的电力负荷预测，只是使用不同的采样方式，这必然会丢失数据真实性。此外很少有文献提到处理大数据下的电力问题，因此本课题将研究考虑多因素下如GDP、大型钢厂的产能、天气气候等多因素下的中长期电力负荷预测并着重提高预测精度，将考虑运用基于量子搜索与粒子群结合的神经网络算法。在数据量大的电力短期负荷预测问题上，提出基于大数据的电力负荷预测，保证精度的情况下运用解决数据量大导致算法运行效率低等问题。

本论文在电力负荷背景下研究改进了神经网络算法，将量子算法和粒子群算法相结合应用到神经网络算法中，使其达到不依赖初始值，收敛速度更快，而且使其更有能力达到全局最优。在此基础上我们还提出将运用分布式的方法，提高算法运行时的效率问题，以达到大数据情况下能有效部署目标。在应用价值上改进的算法能运用在手势识别，图像识别等分类预测的实际应用中，而在数据量大时，运用分布式形式来有效提高运行效率。因而可将其进一步运用到数据量大的预测地点如股票市场，银行等。

## 1.3 国内外研究现状

伴随着当下电力工业的快速发展，电网的优化管理则越来越重要。这一切的先决条件就是能准确预估电力负荷趋势，这也是当今电力系统保持稳定和实现自动化调度的理论依据。

电力负荷预测如果按照预测的期限分类可以分为：长期负荷预测，中期负荷预测，短期负荷预测。长期负荷预测一般是指时间为以年为单位的期限，通过多年度电量、年度电力等指标为内容进行预测。中期负荷预测一般是指时间为月为单位的期限，通常以月度指标进行预测。短期则是以周、天、小时等为单位，来预测某个高峰期的电力负荷。本文将在小数据量时，预测中长期的相应的指标，同时在数据量大时，预测出短期的电荷的相应指标。前者一般使用的方法有类比法和专家估算法等，后者一般可采用灰色模型[8 , 9]，回归分析预测，基于时间序列的预测方式等。但是这些方法的实验结果会受到多变量，非确定等因素影响，因而单一的预测方式不能适应更加复杂的情况，需要能将所有因素综合考虑的组合方式进行建模预测。

随着机器学习算法的蓬勃兴起，电力负荷预测早就从单一变量延伸到了对多因素多变量问题的解决。那种依靠电力工作人员常年积累经验来对未来的电力负荷发展预测估算，早已不适合时代发展的潮流。进入二十一世纪，随着电力交易的更加开放，信息化程度高，电力系统更加复杂化，预测方法需要更加科学，预测的准确度需求更高，这也极大的促进了电力负荷预测理论研究的开展。

当前，国内外研究电力负荷预测的论文比较多，这其中使用的预测模型和分析方法也多种多样，总结起来包括两大类的方法：传统负荷预测，新兴的预测。

### 1.3.1 传统负荷预测

传统的电力负荷预测一般挖掘出电力负荷本身的规律或者电力负荷与一些简单的变量之间的关系，从而运用这些规律对未来进行负荷预测。但这种方式所预测出的结果并不是让人十分满意。概括常见的方法有以下几种：

1. 时间序列法

时间序列法是在七十年代提出来的。通过对电力负荷的历史数据的采集，将电力负荷数据看成以时间为变量的增长序列。因而可以对历史数据的统计进行系统的分析和研究，找出电力负荷随时间变化的关系，以此来建立时间序列模型。最后通过这样一个模型来对未来的某地区的电力负荷进行趋势预测。时间序列是按照历史统计的规律，即认为未来是过去的延续思想。

时间序列法中一般使用的模型有自回归模型，自回归-动平均模型，积累式自回归-动平均模型等。其中最常见的方法Box-Jenkins时间序列法。该方法将负荷序列当成一个随机过程处理，可以应用到上述几个模型进行预测。一般时间序列法模型拟合曲线可以是：直线，二次曲线，修正指数曲线，N阶多项式曲线等。

由于运用时间序列法所需要的历史负荷数据量少，很方便的就能反应出负荷变化，在电网的运行情况良好，不受其他额外因素影响时，对预测的结果能有很好的效果，但是这些都是在没有大的数据波动的情况下所进行的。一般而言，现实生产过程中肯定会受到多方面的因素影响，这又使得其预测的实际结果并不是那么严谨。

1. 回归分析法

回归分析法是一种统计学方法。一般情况下，回归分析需要确定预测值和影响因子之间的关系进而预测出所需值。因为回归模型是对历史数据的重新构造，因而此方法能很好的对历史数据进行拟合，同时由于越久远的数据其实用性就会越低，必然会影响精度。

传统意义上的回归分析，并不能很好的辨别出噪音问题，此时训练出来的模型无法真实反映出变量和实际的预测值之间的关系，因而有较大的局限性。此时，有些专家则提出了一种新型的回归分析方法，此方法将模糊理论与回归分析相结合，对回归系数做模糊化处理，得到了较好的预测结果。

1. 曲线回归法

曲线回归又可称之为趋势外推法。当电力负荷变化与季节因素相关性很小时，则就可以根据已知的历史数据拟合得到一条曲线。该曲线可依据时间变量，因变量，建立这样一个函数。这样就可以通过函数变量在中找到所要预测的值。一般对预测由以下几种处理手段：

(1) 水平趋势预测技术：假定获取到的负荷变化数据为都在期，并且都有水平趋势，则从这组数据出发求得内插值 以及其预测值 。

(2) 线性趋势预测技术：其中包括二次指数平滑法，二次滑动平均法和二阶

自适应系数法。

(3) 多项式趋势预测技术：如果函数足够光滑，可运用多项式模型进行拟合，

从而得出预测值。

曲线回归法的优点在于：适用于数据量少，有效的预测短期负荷。其缺点是：不能随动态的负荷变化进行预测，会增大误差影响结果。

### 1.3.2 新兴负荷预测

1. 人工神经网络

神经网络最早的研究是从40年代开始，有心理学家和数理逻辑学家Pitts首先提出了一个简单的神经网络模型。它是模拟动物神经网络行为特征的一种方法。这种预测方法能够从样本数据中自动学习以前的经验，通过样本数据建立一个规律模型，然后进行预测。由于神经网络具有多个神经元结构，因而它适用于分布式并行处理，显然具有较快的处理速度和较高的容错性。

人工神经网络可自主学习，在学习的过程中进行优化，因而能得到一个较好的结果。但是人工神经网络也有其局限性比如：容易陷入局部极小值，当网络建立的过于庞大时收敛速度会出现缓慢的情况，同时他所需要的样本容量比较大等。

这就出现了以神经网络为主体的改进算法。如文献[1]中将粗糙集理论同BP神经网络的研究结合起来，用粗糙集对影响电力负荷预测的因素进行提取，以此确定模型的输入变量，但这都是短期进行了预测，没有对国家政策影响进行必要的考虑。文献[2]采用了改进的粒子群算法和BP神经网络结合，提出在粒子在计算位置、速度迭代时，粒子数量可以迭代次数增加而增加，并提出粒子可以以相同方向但不同速度进行寻找最优值。

2. 支持向量机

支持向量机是由Corinna Cortes和Vapnik8首先提出的方法，它是一种线性和非线性数据进行分类的方法。支持向量机将向量映射到一个高维空间中，在此空间中建立一个超平面，将两个平行的切面分割超平面，要求让这两个超平面的分割距离最大化。

除了线性SVM 方法外，在线性不可分的数据（也称非线性可分的数据）的分类也可创建出非线性SVM。这种支持向量机可以分辨出输入属性形成的空间的非线性分类边界。得到非线性SVM主要通过：1.使用非线性映射将原始输入变换成高维空间中。2.在映射的空间中寻找超平面。然而在支持向量机不可避免的也有其相应的缺点：例如最快的SVM训练的也很缓慢。但优点还是很多：于其他模型相比不太容易过多拟合，由于复杂的非线性边界建模能力，预测准确度更高等。

1. 小波分析法

小波分析的应用是数学研究成果转向工程学的另一个代表。小波分析源于对原始信号经过变换，进一步用于数据压缩、特征等，从而达到对数据的降维。它运用小波基，提取信号中的指定的时间和频率，使得在时域和频域中都达到比较好的局部化。在进行电力负荷预测时，可以对负荷以时间序列的形式进行小波变换，这样就可以将相应的序列映射到不同的尺标上，映射后的尺标的各个子序列就代表了原序列中不同域频的分量，从而就可以代表负荷序列的周期性。在通过上述操作，得到各个子序列的预测情况，最后对子序列进行还原，完成预测结果。

1. 组合模型法

由于对电力负荷的影响因素多种多样，我们很难从单一模型中建立起对结果的精确预测。这是由于单个模型的或多或少尤其自身的局限性，使得得出的结果难以让人满意，这就进一步的促进了组合模型的发展。在国内外相当多的文献都运用了组合模型的方式进行算法改进。例如可以运用启发式算法（模拟退火算法，粒子群算法，遗传算法等）对神经网络进行迭代次数的优化。这样能综合各种算法的优点，提高预测的精确度和效果。

本文就是运用组合模型对神经网络的进行相应的改造。

## 1.4 论文章节安排

本论文主要分为6章，各章内容安排如下：

第1章 绪论。本章节重点阐述了论文的研究背景，并提出论文的研究目的和意义，同时还对国内外对电力负荷相关工作进行简要介绍，最后对章节安排作出说明。

第2章 相关技术。本章详细阐述了BP神经网络算法，粒子群算法以及量子粒子群算法原理，并简要介绍了各算法的优缺点。与此同时还详细对算法的公式进行了必要的推导。本章还阐述了MapReduce分布式处理框架的原理和执行过程。

第3章 神经网络优化算法。本章首先对选用神经网络算法应用电力负荷预测原因进了说明，同时对BP神经网络在电力负荷预测中的缺点进行了详细的分析。鉴于它的一些缺点从而引入粒子群优化神经网络来提高收敛性和学习的速度。在对文献的分析基础上发现了粒子群优化神经网络的不足，从而进一步提出了以量子粒子群优化神经网络以及一种对该方法进一步优化措施。同时针对改进的量子粒子群算法，在相关函数上与几种算法进行了实验测试，并证明了改进算法的有效性。

第4章 基于分布式神经网络优化算法。本章针对神经网络在处理电力大数据方面上的不足，分析了分布式神经网络优化的必要性。鉴于此本章实现了神经网络并行化以及对其过程进行了解，同时对粒子群优化神经网络分布式实现过程作了重点说明和实现。最后在此基础上提出来量子粒子群优化神经网络分布式步骤过程。

第5章 针对电力负荷方面的数据对本文提出的几种算法进行了实验测试对比。本文首先将改进优化算法应用到中长期电力负荷预测，分别从迭代次数，精度以及寻优结果方面实验说明。然后本文针对短期电力负荷运用MapReduce部署了优化算法，并验证各种算法的收敛速度，计算加速比。

第6章 总结和展望。对全文工作进行了总结，分析在对神经网络改进算法以及分布式实现过程中的不足，同时解剖自己在论文完成过程中的欠缺之处。针对以上的不足之处制定出以后的研究思路。

# 第2章 相关技术介绍

## 2.1 神经网络算法

人工神经网络[11,12]又称为神经网络,它是由大量的神经元组成的一个网络结构,是对人大脑的抽象和模拟。神经网络就是通过这些简单的神经元实现复杂的非线性函数映射，从而使他具有大规模并行处理非线性模型的能力。此外，神经网络不依赖所处理对象的精确数学模型，输入的样本通过训练数据在网络中拟合出合适的模型，通过测试集来测试其精确度，就可以对所需要的结果进行分类预测。

人工神经网络通过改变连接参数来调整整个系统，使它能够适应由于环境的变化所带来的影响，从而实现类似人的学习，归纳，分类，预测等功能。首先需要对被处理数据进行训练学习，建立起合适的神经网络模型如输入输出节点数量等，然后通过训练算法调节各个神经元之间的权值和阈值，给定输入信息，训练后得到输出数据。此时，各个神经元之间的权值和阈值都训练好。最后只要在训练好的模型中输入相应的信号数据，则会在神经网络记忆中返回训练时相应的结果。

### 2.1.1 反向传播神经网络模型

BP神经网络是一种有导师的学习算法。在多层感知器中，学习方式采用反馈传播形式神经网络称为BP网络。BP网络一般具有三层或三层以上结构，包括输入层、隐含层、以及输出层，其中隐含层可能多个。各层之间可实现信息的相互传递。BP神经网络学习过程可分为两个部分：输入层到输出层的正向传播和输出层到输入层的误差反向传播过程。

1. 正向传播

输入层经过隐含层最后到输出层这就是信号的正向传播过程。在信号传播过程中网络的权重是不变的，上一层的状态只会影响下一层的状态变化。如果输出的值满足不了期望输出的值，则转向误差的反向传播。

1. 反向传播

误差值是指网络的实际输出和期望输出之间的差值，误差是由输出端逐渐的向后传播到输入端，这就是误差的反向传播。在误差向后传播的过程中，各层权值由误差反馈调节，通过调节各层权值，使网络中实际输出的值更加接近期望输出的值。

神经网络相关理论已经证明，只要隐含层的节点数量足够，即使是只有1个隐含层也可以以任意精度逼近任何一个非线性函数。相关实验证明，隐含层数如果越多，误差在传递过程也会丢失越多，这样网络泛化性能也就会相应的降低，因此本文在BP神经网络也采用的是三层网络结构。下面是一种常见的BP神经网络模型。



图2.1 BP神经网络传输模型

Fig 2.1 BP neural network model

如图2.1中，其中输入层，隐含层，输出层的节点个数分为为，(i =1,2,….,m)为网络中的输入值，(k=1,2,…,q)为网络中的输出值，为输入层到输出层的权重， 为隐含层到输出层的连接权重。

### 2.1.2 反向传播神经网络推导过程

本文选定的激活函数为Sigmoid函数即

1 输入值向前传播过程：

①隐含层j的输入：

②隐含层j的输出为：

③输出层k的输入：

④输出层k的输出为

⑤从而可以得出输出层神经元k的输出误差为：

⑥对每个样本的总输出误差为：

2 误差向后传播过程：

误差向后传播过程就是对权重更新的一个过程。在计算总输出误差的函数时用到梯度下降法，逐层的改变连接的权值。对于每一个权值w，计算出相应的，然后对权值w进行变更,经过迭代后达到一定精度后使误差达到期望值。

①输出层单元的权值修正过程：

其中

由sigmoid函数的导数为 故可得出：

从上可以得出输出层的权值变化为：

②隐含层单元的权值修正过程：

在网络中隐含层的权值变换，需要考虑到反馈过程中输出层对其权值变化的影响。

### 2.1.3 BP神经网络算法描述过程

下面三层BP神经网络算法步骤：

1. 建立网络，设定输入节点，隐含层节点，输出节点个数，随机设定网络初始化权重和阈值，并设定学习步长，选择隐含层，输出层的激活函数。

2. 输入一组训练样本，将训练样本的属性作为输入层输入，并向网络中输入期望的输出值。

3. 每个训练样本，将其向前传播：输入层到隐含层，隐含层到输出层，最后输出结果值。如果此时的输出值得实际误差值小于指定的误差，则结束训练，反之，则进行下一步过程。

4. 误差的反向传播过程：根据输出层和隐含层之间的权值连接反向传播的结果计算，并更新的值，根据隐含层和输入层之间的权值连接反向传播结果计算出，并更新的值。

### 2.1.4 BP神经网络影响因素分析

BP神经网络优化对样本的采样，网络结构的选择，激活函数选择，初始权值的设置等都会有要求，他们都会对神经网络的性能产生很大影响。

1. 初始权重对神经网络的影响

BP神经网络的训练从初始权值开始，其总是沿着误差函数的下降方向达到极小值。因此一旦我们将初始值确定了，就会确定网络的搜索范围。因此选择合适的初始权值对最后极小值的寻找十分重要。在研究初始值选择的文献中有很多方法如：

1. 随机选取区间为[-1,1]的初始权值。
2. 将两个网络层传输过程的初始权值采用不同的初始权值：输入层到隐含层之间权值连接采用很小的随机数进行初始化，隐含层到输出层采用方法a)中的方式。
3. 将权值初始为一个区间为[a, b]之间的随机数，其中a，b满足以下公式：

其中H为隐含层节点个数。

1. 隐含层个数对神经网络性能影响。

我们知道神经网络的层数越多，分类的精度就会有相应的提高。但是如果层数过多，不但会使网络更加复杂，而且会出现过拟合的现象。研究表明，三层结构的神经网络模型就能够对非线性连续函数进行逼近。如果我们对合适的隐含层网络节点数，则将会进一步降低误差，使得预测的效果更好。隐含层节点个数越少，网络传播的信息就会越少，预测的精度就会有所下降；隐含层节点个数越多，网络就会越复杂，但是学习的时间就会有所增加。

### 2.1.5 BP神经网络常见的优化算法

对BP神经网络的优化常见的是采取措施对收敛速度进行提高或者通过运用智能算法对网络中的权值寻优过程进行必要改进。

一. 收敛速度的提高常用有以下方法：

1. 增加动量项

为了使训练过程中收敛速度加快，提高训练速度，一般的BP神经网络算法都增加了动量项，即

其中且，为动量项。

动量项通过对前一次迭代的权值调整，来影响下一次权值的调整。动量项中包括了历史的经验，当误差面出现了震荡时，动量就会考虑到历史经验对这种震荡起作用，从而提高了训练速度。

2. 自适应调节学习率

学习率也称之为学习步长，一般情况下，很多实验都会将其设定为一个常量，但是在实际过程中，很难确定一个不变的学习步长。如果学习步长太小，就会增加训练的次数；如果学习步长太大就可能增加迭代的次数。因此需要在实验过程进行验证学习率的最佳大小。

二．调节权重的算法

神经网络的训练过程最终可以归结寻找网络中使总误差最小而达到的最优权重问题。因而对最小值的寻优问题，我们可以通过引入智能算法去寻找极值，如遗传算法，模拟退火算法，粒子群算法等。

## 2.2 基本粒子群算法

### 2.2.1 算法起源

Heppner、Reynolds等人通过观察鸟群的行为活动，提出来对鸟群行为的模拟。他们发现鸟类在觅食过程中经常突然改变，其运动轨迹总是不可预测，但总体上总是保持一致性。这些科学家都认为这些行为是基于某种鸟类社会的特殊群体动态学。在早期的模型中仅仅依赖个体间距的操作，也就是群体的各个个体通过努力保持自身的最优结果，从而再从中找出最优的结果。

根据鸟群行为的模拟，Heppner、Reynolds于1995年提出来PSO算法[20,21]。PSO算法不是仅仅的依靠个体自然进化，也需要依靠群体活动对整体进化模拟。也就是说，生物群体中的个体存在个体对个体的影响，个体对群体之间的相互作用，而这些行为体现的是一种生物群体间的信息共享机制。 信息间的共享，正是PSO算法需要模拟的行为。PSO基本思想就是根据此类过程总结而来：将所要求的变量值看成一个粒子，每个粒子在变量的维度以一定的速度进行搜索，通过适应度函数淘汰粒子，粒子根据自己最好的位置和其他粒子的位置，来调整飞行速度和移动位置，最终得出最优解。

### 2.2.2 粒子群算法原理

假设群体为M的粒子群，在D维目标空间搜索，设为，其中，表示粒子在空间的位置， ，，表示第个粒子的速度，。对于每个第粒子在搜索过程中的自身行动轨迹最优值，在整个群体中，所有粒子搜索的全局最优值表示为，其中。

粒子更新方式如下：

其中，；k为迭代的次数；和为正常数的学习因子，分别是为了调节粒子和飞行步长，和分别为均匀分布式在之间的两个随机数。在实验过程中为了控制速度需要将其设定在一个合理的区间，即，同理 。

从( 2.13) 和(2.14)两个式子可知，粒子的速度更新有三个部分组成：第一部分为粒子的上一次的飞行速度，第二部分为粒子当前位置和它历史最好位置之间的距离 。它代表粒子的“认知部分”，可表示粒子在飞行过程中对自身的不断修正，对自己的飞行方向的认识。第三部分为粒子当前位置和所有粒子最好的位置的距离。它代表粒子的“社会部分”。可表示各个粒子在飞行过程中相互间的通信，使自己不断的靠近群体最优位置。

### 2.2.3 粒子群算法实现步骤

1. 初始化。随机产生一组规模为M的粒子群 随机产生其对应的速度,并设定上限的 ,初始化和为相对较大的一个值。

2. 计算出每个粒子的位置的适应值。

3. 对于每个粒子中位置信息，计算出当前每个粒子位置的适应度，保存当前最好的适应度，同时计算出全局最好。

4. 计算每个粒子的更新速度和位置。

5. 如果得到的适应值足够好或者已经迭代到最大次数或者达到设定的期望准确度，则停止更新，否则返回第二步。

### 2.2.4 粒子群算法相关改进

1. 带有惯性因子的粒子群

基本粒子群算法和其他优化算法相比，虽然具有设置参数较少，寻找最优的速度快的优点，但是其还是有很多自身的问题如很容易陷入局部最优，全局寻优能力较差，算法不能保证以概率为1收敛到全局最优等。鉴于此，和为了改善算法在收敛过程中的性能[21]，引入了惯性因子概念即将更新公式改写成如下的方式：

这里的为非负数，称之为惯性因子。它的作用就是平衡全局搜索和局部搜索能力的，当大的时候，全局搜索能力就越强，当小的时候，局部搜索能力就越强。因而选择适当的非常重要。这就是我们常说的标准的PSO算法。

当惯性因子取1时，此时标准PSO算法就是基本PSO算法，因此可认为其就是对基本PSO算法的一种扩展。在对确定惯性因子取值大小时，现今有很多的文献资料对其进行了研究通过大量实验结果表明惯性因子的取值范围有较好的效果。但是我们分析后知道 的取值不可能固定为一个值，这是因为在粒子开始搜索时，算法要保证其在全局搜索能力强，当到一定阶段的时候就需要收紧搜索使其在局部搜索能力强。相关文献中建议惯性因子取值如下：

其中表示区间范围内最大的惯性因子，表示区间范围内的最小惯性因子，t表示当前的迭代的次数，表示最大的迭代次数。

2. 带有收缩因子的粒子群

随着研究的深入，1999年Clerc 将收缩因子的概念引入到该算法中，从而提出了一种带有收缩因子的粒子群优化算法。

其中是关于参数和的函数，可以通过下列函数确定：

相关文献研究表明，带有惯性权重的粒子群在求解复杂函数过程中效果好于带有收缩因子的粒子群，但在求解单模的函数问题时则出现的结果后者好于前者。

1. 动态调整学习因子策略

在粒子群算法中，我们总是希望在初始化粒子初期，粒子对自身有较大的“认知”即寻优范围越大越好，对群体有较小的“认知”即粒子间相互影响越小越好，这是为了粒子群算法在对整个寻优空间能够进行全局搜索，不至于陷入局部极小值。在算法寻优后期，我们希望粒子寻优步长变小，以利于算法更好的收敛一个全局最优值，防止粒子跳出最优值。因此相关文献中对学习因子、进行了线性递减的处理，实验证明其寻优效果在一定程度上可以提高。

## 2.3 量子行为的粒子群算法

在粒子群中要使粒子达到最优性，则需要粒子具有聚集性和分散性。聚集性指个体总是倾向学习个体中最优的结果。分散性指粒子在学习过程中能够尽可能分散在各个位置，减少聚集性。但基本粒子群算法中，粒子总是按特定的轨迹运行，这样使得其寻找的区间总是有限的，并不能在覆盖到整个空间的可行解内，这就使得粒子群算法并不能保证能收敛到全局最优解。另外，在Clerc等在研究粒子群算法后认为，每个粒子收敛过程总是在个体局部最优位置和全局最优位置的空间进行搜索，并且根据它的速度迭代公式可知，粒子的搜索空间上是一种线性搜索过程。但在实际生物群体活动过程中，粒子的方向是一个不能确定的量，如果只是用一个线性的速度方向去描述粒子进化过程往往是不够的。

在此基础上，2004年Sun等人将标准粒子群算法同物理基本理论量子结合[]。在量子理论中由于量子的叠加性[26,28]，使得粒子位置产生了很大的不确定性，量子的集聚态，使得粒子能够收敛。在根据粒子群的基本特性，并在受到物理学的量子理论的影响后，sun等在研究了粒子收敛性后，从量子学角度提出了新的PSO模型。这种模型以DELTA势阱为基础，把粒子群中每个粒子看成具有量子特性行为，据此提出了具有量子行为的粒子群算法(QPSO)。

根据文献[9]中基本粒子收敛行为分析得出:粒子在收敛过程中以吸引。 其坐标为：

其中 ，

### 2.3.1 QPSO模型建立

在量子系统中，粒子的位置和速度是不能同时确定的。量子空间的粒子状态是由波函数 来描述。波函数的平方就是粒子在空间中某点出现的概率：

其中为概率密度函数。

在量子空间中粒子运动的动力学方程就是方程:

其中是哈密顿算子，称为普朗克常数。可表示为以下形式：

M是粒子的质量，是粒子所在的势场。

假设粒子群系统是一个量子空间的系统，每个粒子都具有量子行为，并且可以通过波函数描述其状态。为了使粒子不发散，使其聚集在他们的局部点。简单的考虑一维空间中粒子的情况。P点作为势的中心，粒子的势能势阱可表示为：

粒子在势阱中的方程归一化波函数为：

### 2.3.2 量子位置

根据量子力学的原理，粒子的位置和速度是不能同时确定下来的，但只要知道粒子的确切位置就可以计算出目标函数值。通常可以通过蒙特卡罗随机模拟的方式来测量粒子的位置，这种方法相当于对粒子瞬时运动过程中，进行拍照，在那一瞬间，粒子的位置就可以被确定下来。

粒子在某一点的概率分布函数为：

通过蒙特卡罗逆变换，设是区间上的均匀分布随机数，则用

联合两式可得出

其中是区间的均匀分布随机数，L是粒子位置概率密度函数特征长度。一般L取。

从上可以确定出量子的位置表达式为：

其中表示所有粒子当前的最优解，为参数被称为收缩-扩展因子，一般对收敛速度进行控制。

### 2.3.3 量子粒子群的特点

1. 量子粒子群算法中量子系统是一个满足了状态叠加性原理的复杂非线性系统，它比标准PSO粒子更加能够模拟出粒子运行状态变化。

2. 在QPSO算法中，各个粒子具有量子行为，没有一个确定的运动轨迹，它能以某一个确定的概率出现在搜索的任何位置，而在PSO中粒子总是速度的轨迹进行运动，这就很容易限制了搜索区域，使得PSO搜索的空间没有QPSO大，从而可能搜索不到某一区域的全局最优值。但是QPSO中粒子的位置可以远离局部最优值的吸引，从而可以到达当前种群全局最优值的位置。

综上所述具有量子行为的粒子中，将原来的通过改变粒子速度来改变粒子的位置方式，变换成只要粒子坍塌就能确定某一时刻的粒子位置，从而确定最优值的方法。这种方式能减少由于中大量参数带来的调参困难。同时由于量子的不确定性使得其在位置搜索过程中避免了PSO搜索只限于特定空间的问题。

## 2.4 分布式平台

### 2.4.1

是由创始人 创建，是 下的一个开源项目。受到了 开发的 和 的启发。是由计算和存储组成。 分布式平台共有以下四个特点：

1. 高可靠性。 默认冗余存储写入的数据，会自动的维护数据的多份复制。

2. 高扩展性。 是在可用的计算机集簇间分配数据并完成计算任务的，这些集簇可以方便地扩展到数以千计的节点中。

3. 高效性。平台作为一种分布式的架构，能够快速高效的处理大规模海量数据。

4. 高容错性。为了避免数据的丢失，所以能够对数据自动的保存多个副本，并且能够重新分配失败的任务。

### 2.4.2

是 项目的核心子项目，是分布式计算中数据存储管理的基础，是基于流数据模式访问和处理超大文件的需求而开发的，可以运行于廉价的商用服务器上。它所具有的高容错、高可靠性、高可扩展性、高获得性、高吞吐率等特征为海量数据提供了不怕故障的存储，为超大数据集的应用处理带来了很多便利。同时分布式文件系统像目前通用的分布式文件系统一样对外开放自己的操作接口，因而使得用户可以很方便的进行文件的相关操作诸如移动，删除，创建等。图2.2 是的体系结构图。



图2.2分布式文件系统体系结构

Fig. 2.2 The Structure of HDFS

虽然在处理海量数据时具有很多其他的分布式文件系统无法比拟的优势，但是分布式文件系统依然具有不少缺点。主要体现在不适合低延迟数据的访问以及对于大量的小文件不能够进行高效的存储，而且不能够支持多个用户对文件进行任意的修改和写入操作。上的文件被划分为64MB 大小的多个分块，作为独立的存储单元。但与其他文件系统不同的是，中小于一个块大小的文件不会占据整个块的空间。的块比磁盘块大，这样做是为了最小化寻址开销。

### 2.4.3 编程模型

是由公司开发的一个针对大规模海量数据处理的分布式编程模型。实现了两个功能。 把一个函数应用于集合中的所有成员，然后返回一个基于这个处理的结果集。而是把从两个或更多，通过多个线程，进程或者独立系统并行执行处理的结果集进行分类和归纳。和 两个函数可能会并行运行，即使不是在系统的同一时刻。图2.3 是的执行流程图。



图2.3 MapReduce计算过程

Fig 2.3 The executive procedure of MapReduce

一个 作业（）通常会把输入的数据集切分为若干独立的数据块，由任务（）以完全并行的方式处理它们。框架会对的输出先进行排序，然后把结果输入给任务。通常作业的输入和输出都会被存储在文件系统中。整个框架负责任务的调度和监控，以及重新执行已经失败的任务。框架由一个单独的和每个集群节点一个 共同组成。负责调度构成一个作业的所有任务，这些任务分布在不同的上，监控它们的执行，重新执行已经失败的任务。而仅负责执行由指派的任务。

## 2.5

本章主要介绍关于本论文中需要用到的相关算法，包括BP神经网络、粒子群算法以及量子粒子群算法。首先介绍了BP神经网络模型，并分析了影响其性能的因素，同时还简要说明了相关文献的对其的改进。然后介绍了粒子群算法，其中主要对粒子群算法的原理以及相关改进做了描述。本章最后介绍了量子粒子群算法，重点在对粒子模拟量子行为的推到过程，以及引入量子后粒子特点进行了说明。

# 第3章 量子粒子群神经网络优化电力负荷预测

神经网络由于有很强的非线性拟合能力，可以映射任意复杂的非线性关系，同时它还具有很强的鲁棒性等优点，因此它非常适合对电力负荷进行预测。于此同时在中长期负荷预测受到很多因素的影响如：天气变化，国家宏观调控，市场消耗的煤炭量等，这就需要我们建立好更合适的模型对电力负荷进行预测。优化BP神经网络算法以及对多种算法进行组合是算法改进的重要手段。本章根据电力负荷相关特点利用多种算法的组合对神经网络进行必要的优化。

本章首先将会介绍传统的神经网络算法在中长期电力负荷预测的应用所存在的相关问题，以及基于PSO改进的神经网络算法所存在的缺陷不足，接下来本章会提出量子行为的粒子群算法优化神经网络在电力负荷预测的应用，最后论文中将对量子粒子群进行必要的改进再结合BP神经网络。

## 3.1 神经网络在电力负荷预测中的缺点

BP神经网络算法是以梯度下降法为基础的，而它具有线性收敛速度以至于学习训练次数过多并且训练过程过长等问题，同时收敛速度过慢，也是BP神经网络的一个比较大的缺点之一。由于神经网络形成一个多维的空间，则很可能出现在求解最优值时出现局部最优的情况。在此可以总结出BP神经网络在电力负荷预测出现的问题：

1. 由于电力负荷可能受到多种因素影响，如天气，煤炭消耗，以及各种国家宏观调控等因素的影响。这就使得在进行构建BP网络时需要构建多个网络层和各层的多个节点。BP神经网络学习算法在反馈调节权值和阈值过程中是以负梯度下降方式来变化，所以当误差达到一定值时，这种调节就会停止。但是在预测电力负荷时，构建的神经网络计算最小误差时形成的最优函数，可能存在多个局部极小点的超曲面。这就使得BP神经网络在训练时只能收敛到某个局部最小值的情况。
2. 在预测电力负荷时，运用BP神经网络进行迭代训练一般采用的增量训练法，这种方法需要将所有的样本的输入值放入网络中，训练后产生的全部误差再反馈到隐含层和输入层。此时再进行正向反馈，一直像这样迭代。在研究神经网络迭代过程中发现，如果想要网络稳定，需要大量的迭代次数，这就会出现这样的问题：1.使网络稳定，迭代次数多。2.向前传播过程中如果有多个隐含层时，迭代缓慢。
3. 在运用BP神经网络预测电力负荷时，预测出来的值为连续值，而非常见的用于分类的离散值。在很多的文献[16,19]中处理这种连续值时采用的是输出层只有一个节点的方法，但是这将很大程度上降低网络的复杂度，从而对预测结果的精确度有很大的影响。鉴于此种情况本文将在实验部分对连续值进行相关处理，让BP神经网络预测连续值时也能存在多个输出节点，增加网络的复杂度，从而使拟合的效果更佳好。

## 3.2 基于PSO优化神经网络电力负荷预测

### 3.2.1 粒子群优化神经网络的可行性

在进行电力负荷预测时，运用传统的BP神经网络减少预测误差时通常使用梯度下降法来调节权重和阈值，正如第二章介绍的推导公式计算网络中各层的权值优化，但这种方法很容易出现在上述3.1节的问题。针对这些出现的问题，国内外专家做了大量的研究，提出了各种改进的方法，比如采用自适应的调参，运用启发式的算法优化BP神经网络等。粒子群优化算法实质上是一种并行的全局性随机搜索算法，它比遗传算法更加简单的同时搜索能力更加快速等特点。而对神经网络的优化实质上是对误差函数寻优的优化，这就为两者之间的结合提供了一个很好的思路。粒子群和神经网络的结合[22 , 25]主要有以下两种方式：1.利用粒子群的全局搜索能力来优化神经网络的拓扑结构、连接权值，将粒子群算法全局寻优能力和神经网络算法的局部寻优能力相结合，从而提高了网络的学习能力和整体的搜索效果。2.将神经网络结合粒子群算法，利用神经网络学习能力改进粒子群算法的优化性能，从而提高粒子群的收敛速度，同时还能减少计算量。

综上所述，神经网络和粒子群虽然各自有各自的特点，但两者并不冲突，反而互补能力很强，这就为他们结合一起提供可能。本节将考虑利用粒子群相关算法与神经网络结合用来进行电力负荷的预测。

### 3.2.2 粒子群神经网络关键问题

由于PSO是一种优化算法，利用PSO算法进行网络的优化实质上就是一个寻优过程，更加具体来说就是进行向前网络的计算，而不需要进行向后网络反馈。这其中涉及到了两个主要问题就是粒子与网络中的权值的关系以及粒子群中适应度怎么构造的问题。

PSO优化神经网络的关键点在于以下两点：

1建立起PSO粒子个体的表示与BP神经网络连接权值之间的映射关系。也就是粒子群中每个个体怎样表示成求解BP神经网络优化函数的权值。

2将神经网络中的均方误差作为粒子群中的适应函数，通过PSO算法优化过程求解全局位置。

了解了这两个关键性问题，我们就可以设计出相应的映射关系与适应函数。

如图2.1中所示的三层BP神经网络结构，神经网络的输入层节点个数为M,假设隐含层节点个数为N, 输出层节点个数为Q。粒子群对神经网络的优化主要体现在同时优化输入层到隐含层，隐含层到输出层之间的权重，他们分别为 和，从而可以使用相应的编码规则对粒子群个体与权重更新产生映射。则需要的每个粒子维度D为

其中表示为输入层中存在的权值+阈值个数，同样表示为隐含层到输出层中的权值+阈值个数。

另一方面对适应度函数的设计。粒子群优化神经网络对权值的更新本身就是一个反复迭代的过程。通常为保证实现的算法的评估准确程度，往往将数据集拆分成两个部分，训练集和测试集。评估准确性即个体在群体中的优劣程度就需要适应度。在PSO优化神经网络的适应度如下：

其中n 为训练样例的个数，为第样例的实际值，对应的预测值。从表达式上看，适应度越小则说明预测出来的误差就越小。

### 3.2.3 粒子群神经网络算法步骤

根据上文的分析确定了粒子群的种群个体以及适应度的计算方式，利用粒子群算法优化神经网络就成为了可能，具体步骤为以下所述：

1. 初始化神经网络结构。这一部分包括神经网络的输入层，输出层个数，在处理电力负荷预测时选取好相关属性作为输入，并归一化相关数据，本文选取3个节点作为输出；设置其他参数如训练样本，测试样本，以及迭代次数等；根据相关的经验公式如定理计算可得。
2. 初始化粒子群维度D，种群规模M以及在一定范围内的随机初始化每个粒子的初始位置，同时设置相关的其他参数如学习因子，位置的上下限，最大迭代次数等。
3. 计算出每个粒子的适应值，我们这里的适应度就是BP神经网络的总体误差即网络中向前传播的输出值和实际值的差值由公式表示。
4. 如果更新后的粒子的适应度比之前存在的适应度好，则我们就可以将粒子当前的位置替换；同理如果更新后的粒子适应度 比全局的的适应值更优，我们就可以将该粒子位置代替全局。
5. 更新每个粒子的位置 通过公式(2.13)和(2.14)，并检查位置是否超过设定的上下限。如果越界则可取其上下限的值。计算每个粒子的适应度。
6. 若运行迭代次数达到粒子群设置的最大迭代次数时，算法停止并输出此时的和相对应的全局最优值 ，否则返回4 。
7. 将粒子群得到的全局最优位置传递给神经网络作为其调整的权值，从而建立起BP神经网络模型，通过反归一化就可得出预测的值。

算法流程图如下：



图3.1 PSO-BP神经网络流程图

Fig 3.1 The process of PSO-BP neural network

综上，粒子群对神经网络的优化在于将神经网络的反馈调节运用粒子群寻优过程代替。这样做的好处是减少各个层之间的调节计算，因为每次反馈调节都要经过输出层到隐含层，再由隐含层到输入层，当网络节点个数比较大时，这种反馈调节计算过程更新速度就非常的慢，如此的话必然将降低网络预测的整体的预测速度，这也就是为什么单纯的神经网络收敛缓慢的原因。粒子群算法通过对误差的整体寻优，避免了上述不断反馈寻优过程，它能通过自身的简单更新位置方式，找到这样的一个使误差最小化的位置，从而建立起网络模型，使得寻优过程速度加快。

## 3.3 基于量子行为的粒子群优化神经网络电力负荷预测

基于粒子群的神经网络优化比基本的神经网络具有更好的性能。前文已经对PSO优化神经网络对比单纯神经网络以及其中出现的困难进行了详细分析，同时对前者优化的性能提升已有大量的文献进行了相关实验，这里不做赘述。随着对粒子群优化算法的深入研究，目前主要存在两种不同类型的量子粒子群算法。一种以量子位概率幅为基础的量子粒子群算法(Quantumn Particle Swarm Evolutionary Algorithm），另一种是江南大学孙俊等人提出的基于量子力学波函数的量子粒子群算法[27]（在第二章已介绍）。QPSO算法是以量子delta势阱为基础，并波函数的平方表达每个粒子出现在某一位置的概率值，然后利用蒙特卡罗随机模拟的方式进行测量，将量子状态坍塌到经典状态下，从而描述粒子在经典状态下的位置。

在相关文献[28]中的实验结果表明QPSO不论在收敛性能上，而且在收敛速度上更加优于QPSEA。因此本文选择使用QPSO粒子群结合神经网络进行电力负荷的预测。

鉴于此，本文提出了一种基于改进的量子粒子群优化神经网络应用电力负荷算法。该算法利用了QPSO优化神经网络的参数，具有更好的寻优效果。

### 3.3.1 量子粒子群对神经网络优化

量子粒子群的提出是对粒子群的进一步的优化。具有量子行为的粒子群克服了粒子群本身的一些缺点，这样在结合神经网络时对模型优化更加准确。在第二章我们已经分析了量子粒子群相对于粒子群算法本身的优点。量子粒子群算法解决了粒子群算法如下问题：

1. 参数过多问题。

我们知道粒子群的更新方式由位置和速度决定即：

在初始化过程中我们必须首先初始化像等相应的参数，虽然在有些文献中给出了建议性值，但是对不同问题不同背景的应用，参数的调试需要大量的时间，这样也使得相应的组合性算法即对神经网络优化增加了更多的困难，而利用量子行为粒子群算法则有效的解决了这一问题。它只需要对位置公式进行更新进行操作，并且只需对一个参数进行初始化。



图3.2 粒子等待问题

Fig3.2 Particle waiting problem

1. 粒子等待效应问题。

假设我们在初始化随机粒子的时候产出的粒子不是那么分散，还有一种可能就是在粒子群迭代达到一定阶段时只有个别的粒子远离了集聚在一起的局部群体。如图3.2情况，大部分粒子都集中在附近，在远离的的少量粒子，我们称之为落后粒子。在粒子群算法中，对于附近的粒子，粒子位置很快的就会收敛到，而落后的粒子由于离较远，会以较慢的速度向运动。而我们知道此时并不是全局的极值，而是远离它的地方，此时寻找最优的极值的责任就会让落后粒子承担。这样粒子群算法全局搜索能力必然就会降低，容易陷入早熟的状态。如果我们改用量子行为粒子群算法的更新即(2.29)式子就可以新的发现。

在对应的量子粒子群算法更新式子(2.29)中，由于有落后的粒子，计算平均最好位置 后得出的位置更新，会将聚集在一起的粒子拉离。这样就会使聚集在的粒子分散开来，在相对较大范围搜索。这个过程中就会使所有粒子都参与全局搜索，等到落后粒子趋向时，他们就会在更小的范围进行搜索，这样就会使全局搜索能力提高。

### 3.3.2 量子粒子群神经网络算法步骤

根据对上述量子粒子群算法的分析可以，对于优化了更新步骤的QPSO算法与神经网络算法结合的步骤如下：

1. 在初始化神经网络结构后，初始化粒子的种群规模M, 粒子维度D等并初始化 以及。
2. 输入训练用例后，对相关算法进行不断迭代更新直至满足迭代次数或者精度要求
3. 通过下列公式对位置进行更新操作

其中为粒子的平均位置，t为迭代次数，为(0,1)之间的随机数,为参数被称为收缩-扩展因子，它就是量子中的唯一参数。P可由下列公式求得：

1. 通过适应函数计算每个粒子的适应值，并更新 以及。
2. 取出迭代后的最佳作为神经网络的权值进行预测。
3. 利用测试集进行算法测试。

## 3.4 基于改进的量子粒子群优化神经网络电力预测

前两节中我们不仅分析了粒子群与神经网络的结合，同时我们还分析了量子粒子群与神经网络的结合。相对于PSO，在利用QPSO寻优对比过程中发现其更能确保全局性，这是因为它更好的综合局部和全局之间的平衡点，从而达到提高神经网络预测性能的目标。这样的QPSO-BP算法将具有更好的学习能力和泛化能力。但是在实验中发现QPSO算法也有些缺点，它容易早熟收敛，因此接下来的工作重点在于保证QPSO能否有效收敛，同时采取一定措施使其避免早熟。本节的内容通过提高QPSO算法的寻优能力，进一步提高BP神经网络的学习能力和寻优能力。

### 3.4.1 改进的量子粒子群神经网络参数优化

利用QPSO算法进行对BP神经网络寻优，可以提高BP神经网络的预测性能。同时QPSO算法和PSO算法一样，都是一种智能搜索算法。但也存在这样两个问题：1。如何在搜索过程中保证收敛。2.在收敛过程中怎样保证算法不陷入早熟问题。这两个问题的解决对于提高BP神经网络的预测学习能力和泛化能力都有很重要的意义。在此本文将以前人关于粒子群收敛性为基础，同时引入量子粒子群的粒子均值思想对早熟进行必要的改进，从而提出一个带有收缩因子X的XQPSO算法优化BP神经网络,并且通过仿真实验证明出带有收缩因子X的XQPSO在寻优以及在迭代次数上更加优于前文提及的算法，从而也就更能说明XQPSO在优化BP神经网络性能更加突出。

1. 算法收敛性

在前文提及到为保证PSO算法收敛性， 将收缩因子的概念引入到PSO算法中,他本身目的就是增强算法的收敛性。为此本文以此为启发，提出将收缩因子应用到量子粒子群中。在 算法中，他将收敛因子引入到了粒子的更新速度上，使得个体在搜索能力和收敛速度上都有一定的提高。

其中 ,研究表明的取值对算法的收敛性有着重要的影响。在一些文献中的作者提出了将的取值范围限定为时PSO算法会很快的收敛，通常取值为4.1，从而值为0.729，。

根据对于PSO算法收敛性分析，本文提出线性递减的收缩因子来调整QPSO中粒子位置的更新，使粒子在迭代过程中能够不断的吸引粒子更新，本文重新定义的线性递减收缩因子为：

则粒子的更新公式如下：

由此粒子就可以通过对进行调整使其收敛性增强。

本文提出线性递减的收缩因子目的是为了解决收敛性问题。相关文献在PSO算法中的最好取值在一定范围，但是实际过程中的取值不应该是不变的一个定值，因而本文为了更好的模拟出生物本身寻优的特性，提出了线性递减的收缩因子。这样做的原理在于：由于具有量子特性的粒子本身具有位置不确定性，这就解决了粒子在寻优过程中可以分布在不同位置的问题即可以让粒子尽可能分散在各个位置；其次对于收敛过程，我们总是希望粒子在初期能够尽可能大空间搜索，而在将要接近找到全局最优值时，粒子的搜索应该变得缓慢，这是为了防止粒子跳过全局最优值。

2．算法的早熟

量子粒子群优化算法在寻找最优解过程中，如果某个粒子个体发现了当前的最优位置，其他的量子粒子个体就会迅速的向这个量子粒子位置靠近。如果当粒子的位置的适应值是一个局部最优值，那么整个量子粒子群就会积聚在这个局部域中停止不前，这就是优化算法中常常碰见的“早熟现象”。粒子群在迭代过程中收敛总是在局部最优值中寻找，以至于总是不能在全局最优解中寻找到最优值的现象我们称之为早熟。这里说的早熟本质上就是当粒子群陷入了局部最优时，总是不能跳出局部最优解，而是陷入在其中。鉴于此我们需要跳出局部，分散寻优的措施。因此在量子粒子群中，本文引入这样一个判断早熟的标准：

其中 分别为粒子的当前位置，粒子自身的最好位置，粒子全局最好位置，迭代次数。当粒子满足上述条件时我们就可称之为早熟即如果对于粒子群的所有个体在迭代初期，就出现了对于其个体本身最好位置与个体位置只差小于一个阈值，并且全局粒子的最好位置与个体位置同样小于一个阈值，那么就可认为其已经陷入早熟。如果量子粒子群优化神经网络寻优过程中出现上述的现象，我们将进行以下早熟收敛处理即粒子的分散性。

3. 粒子的分散性。

前一小节我们已经运用了更好的方式保证了量子粒子群粒子能够按照自己的调整收敛。这里我们就需要保证粒子寻优过程中种群的多样性。这里的种群多样性指的是一种能跳出聚集一起的局部粒子范围从而有效地减少粒子等待效应。

其中为粒子的位置，n为粒子的个数。这样只要出现了粒子等待效应，相对远离局部聚集区的粒子就会有个拉伸的机制将聚集的粒子拉离，从而达到跳出的目的。

### 3.4.2 几种粒子群优化算法对比仿真实验

为了验证本文提出改进的量子粒子群XPSO算法在克服收敛早熟等问题的性能优势，本节将对标准PSO算法，QPSO算法以及XQPSO算法进行测试工作，从而为后续XQPSO证明BP神经网络寻优对比其他算法在性能优势打下基础。

本实验中的寻优目标函数为，其最小值为0，在实验中设定其定义域。目标函数数学表达式如下：

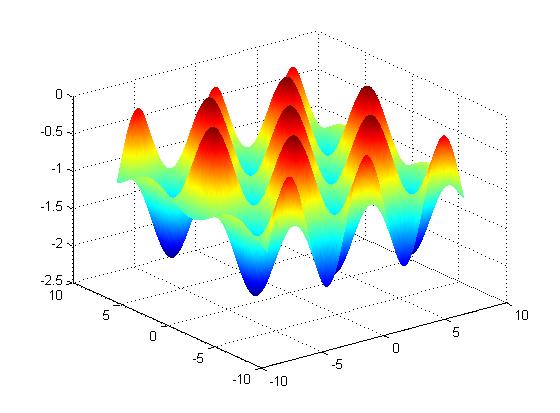


图3.3 三维

Fig 3.3 Griewank function

如图所示为三维中 函数，其在定义域内存在多个极值点。对全局极值点来说，其最小值为0，在局部位置，它有其他的极值。它的这种特性很好的满足了我们所要求解的要求即使用多种粒子群优化算法对全局最优值进行寻优。实验中，如果XQPSO不能从定义域为函数中找到极值0，或者说其在其他性能上没有能够有效提高，则认为其性能没有PSO,QPSO算法成功。

本次试验在求解函数统计次数为50，变量中i的取值为8，各种算法的粒子种群大小为20，其他参数按照本文在实验部分提及到的进行。结束条件：1. 最大迭代次数为10000。2.目标函数值。

表3.1 PSO,QPSO,XQPSO算法性能测试

Table 3.1 PSO ,QPSO,XQPSO Algorithm performance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 相关算法 | 最小适应度 | 最大适应度 | 50次能求出最优值得次数 | 求得最优值时迭代次数 | 求得最优值平均迭代次数 |
| PSO | 0.00075 | 0.1569 | 30 | 1900 | 1932 |
| QPSO | 0 | 0.324 | 27 | 103 | 108 |
| XQPSO | 0 | 0.029 | 45 | 39 | 62 |

从表中4.1可以看出，上述实验的三种算法均不能保证每次都能找出最优值，PSO算法在50次实验中只是求出了最接近的值即最小适应度，没有找出极值0，在QPSO和XQPSO算法中则有更大的概率找到了最优值，但是QPSO稳定性并没有XQPSO那么好。在运行50次XQPSO算法，其找出最优值的次数基本上是运行一次就能找出一次最优值结果，这说明XQPSO算法能够有效的避免陷入早熟问题。求得取最优值迭代次数的指标上，XQPSO在一共10000次迭代次数中只需要迭代69次即可求出最优值，这远远低于PSO迭代次数，这说明XQPSO算法加入X收缩因子起了很重要的作用即能加速收敛的作用。在10000次迭代中QPSO与XQPSO算法都能找到极值点0，说明了QPSO算法在搜索能力上更加优异，同时能够跳出局部最优，拉伸了种群中粒子的位置。

通过对三种算法对函数进行的仿真实验，我们可以得出这样的结论，QPSO较PSO更能找出最优值，这是因为其更好的对生物寻优过程进行了模拟，拥有更好的搜索能力。XQPSO比QPSO在迭代次数上更具有优势，这是因为其引入了公式X惯性因子使其迭代次数上更早的找到最优值，同时在寻优次数上更加优异的可能性在于我们引入了早熟的判定条件，并对其进行了粒子进行了分散处理。

## 3.5 本章小结

本章首先分析了神经网络在电力负荷预测方面的缺点，然后提出了基于PSO优化神经网络在电力负荷方面应用，其中我们在预测连续值进行的必要的处理。同时在对PSO优化神经网络相关文献进行了研究发现其仍然在某些方面上的不足，进而提出了使用量子粒子群优化神经网络，并在此基础上调整了量子行为的收敛性并对其进行必要改进。最后在上文提及的算法基础上，进行了PSO,QPSO，以及XQSPO算法的性能测试。通过比较，验证了XQPSO在处理早熟以及收敛性上更具有优势，从而更好的结合BP神经网络进行后续实验。

# 第4章 分布式神经网络优化算法的电力负荷预测

随着电力智能化程度的不断提高，电力系统数据的海量化出现。这主要由于技术的发展，使得对数据的采样过程发生了变化，在短期负荷预测中电力负荷以秒、分钟或者小时为单位形式出现的越来越普遍，这样很容易就会形成大量数据的累计效应。于此同时在研究分析影响电力负荷预测方面的因素多样化，这使得预测范围更加广泛，预测的效果更加精准，这就形成了我们常说的数据的多维化因素。但是在处理电力负荷数据海量化，多维化时，我们常常不堪大数据的困扰，通常做法是经过抽样的方式采集数据再运用适应的算法进行训练，这种做法虽然减少了数据处理的量，但是却也是牺牲了预测的准确性为代价。

鉴于此种问题，本文将并行化处理电力负荷数据，提出基于分布式神经网络优化算法对电力负荷预测，不仅可以缩短训练所需要的时间，而且还能提高电力负荷的准确度。

## 4.1 分布式神经网络优化研究的必要性分析

电力系统短期电力负荷是以秒、分钟、小时或者天为单位形成的数据，它的一个显著的特点就是数据量大。而对电力负荷预测则需要我们认真考虑影响其变化因素，这就会形成多维度的因素。正如我们前面提及的神经网络进行预测时更适合那种数据量大的数据集，这样预测的准确度更加高。然而海量化，多维度的数据带来的将是以下问题：

1. 训练数据过程缓慢。神经网络在学习过程中需要将数据一分为二：训练数据，测试数据。训练过程中，神经网络需要将所有样例输入求取其误差值，然后通过反馈求取各层之间的误差值。我们考虑的另一个问题在于，一般情况下神经网络如果要真正的收敛的话，一般情况下需要迭代几千次。当训练数据过大的时候，每迭代一次都会去计算总体的所有样本的误差。相对于上千次的迭代的计算过程将是一个庞大的计算量。因此在当面对海量化数据时，单机训练数据已经很难得到相应要求，很容易导致训练时间过长。
2. 网络模型过于复杂。在进行电力负荷预测时需要考虑的影响电力负荷变化因素很多，这样才能使预测的准确度更好。但是这又会出现这样的一个问题：考虑影响因素越多，那么维度就会越大，输入层的节点数就会越多，这势必会带动其它层的节点个数的增加。这样一个问题的出现就会使得整个网络层的模型节点个数的增加，导致神经网络不管是向前传播还是向后反馈时计算过程大大增加。而这又反过来形成训练数据量过多，带来训练过于缓慢的问题。

解决这两个问题的关键点在于对训练数据的有效划分，本文将以上述问题为切入点，基于当前主流分布式框架的平台之上，解决电力负荷预测中出现的海量化，多维化数据集问题，并设计出相关的分布式粒子群优化神经网络优化算法。

## 4.2 分布式BP神经网络算法步骤

Chu et al在论文[40]中提出了基于MapReduce实现后向传播的神经网络对数据进行分类并且做了一些相关实验。但是他并没有将此方法应用到处理电力负荷预测当中，他提出的解决问题是一种分类的问题，也并没有解决预测结果为连续值问题。本节将实现分布式神经网络在电力负荷预测上的算法。

实现分布式BP神经网络主要思想在于将训练分为阶段和阶段。实现分布式的BP一般采用的是增量的方式即将所有样本一起同时输入处理，计算出整体误差，然后再通过改变权值，最后在Reduce阶段对所有权值进行一个规约。具体步骤如下：

1. 构建BP神经网络，初始化BP神经网络的输入层隐含层以及输出层。随机初始化出BP神经网络中的权值和阈值。
2. 读取HDFS中训练样本信息，并将其转化为的键值对，这就是Map函数的输入值。Map函数的作用在于对每个训练样本进行向前传播运算，计算出最后的误差值,通过误差值向后传播得到每个权值对应的改变量。对于每次的权值以及其改变量，在Map函数中输出。
3. 接着就是Reduce过程，Reduce函数将相同的权值传递给同一个Reduce函数处理，其将所有权值更新累加，求取平均值。将这个平均值最为最新的权值调整，写入文件中，作为下一次迭代的输入。
4. 重复Map和Reduce阶段，直到算法迭代到终止。

执行流程图如下：



图4.1 分布式BP神经网络

Fig 4.1 The Process of Distributed BP neural network

1. 分布式神经网络Map函数。

在Map函数中输入键值对，在伪代码function1中，key表示输入的训练样本一行的位移量，value表示样本值信息。Map函数处理过程中首先计算该样本的网络误差值即向前传播过程中误差，然后向后传播计算该样本的误差改变量,这样对于多个Map函数就可以处理尽可能多的数据行。输出的键值对以输出。

2. 分布式神经网络Reduce函数。

在Reduce函数中输入的键值对，在伪代码function2中，相同的key 将其集中一起，对value取每次权重的变化量取平均值，这样就算是以此对网络权重的更新。然后根据这次的更新操作，继续训练。

分布式BP神经网络Map过程

|  |
| --- |
| Function1: 输入：<key1,value1> ;  输出：<key2,value2>; |
| //向前传播计算出总体误差值,通过从获取的权值和样本的值。  //向后传播计算权重的改变量。  Output(  } |

分布式BP神经网络Reduce过程

|  |
| --- |
| Function2: 输入：<key2,value2> ;  输出：<key3,value3>; |
| Sun=0;  value=0;    Sum+=value;  count+=count+1; |

在主函数设计按如下方式进行：

在job中运行Map和Reduce函数在并计算得出reduce的输出后，做一次总体的权值更新操作即对网络中的权值进行一次更新，如果训练样本集网络总误差小于期望的最小精度，则再次进行function(1)和function(2),否则运行结束。

## 4.3 分布式神经网络优化算法实现

前面一节我们已经实现了对分布式BP神经网络的算法。目的是为了解决神经网络中的计算速度慢的问题，但是并没有解决上文中提及的陷入局部最优的问题。鉴于次本文提出分布式PSO-BP神经网络算法、分布式QPSO-BP神经网络算法以及XQPSO-BP神经网络算法应用于电力负荷预测，用来优化分布式神经网络减少其陷入局部最优值的可能性。

### 4.3.1 分布式BP神经网络优化算法关键问题

经过上文的分析，探究了分布式神经网络在解决电力负荷预测的必要性。但是真正要实现分布式PSO-BP神经网络需要解决以下问题：

1. 实现分布式计算的Map输入输出和Reduce输入输出问题。要实现分布式PSO-BP神经网络关键点之一，平台下怎么考虑PSO和BP神经网络的MapReduce计算过程。PSO算法中种群在更新自身的最优值同时还需要更新全体最优值，这一特性很符合Map和Reduce处理过程。在Map 函数输出可以对每个种群个体进行更新，并输入到Reduce函数中集中处理种群，再由Reduce输出最优的。BP神经网络中在设计成分布式的时候，增量输入更加符合分布式的部署。Map函数输出为每个权值以及权值的变化量，并输入到Reduce函数中对每个权值的变量均值处理，最后输出权值更新后的值。
2. 在分布式下BP神经网络和粒子群算法的结合问题。在非分布式下粒子群优化神经网络算法是一种串行化过程，分布式下PSO结合BP神经网络我们需要考虑更多的是处理数据的能力上，以及计算过程。因此我们可以将BP神经网络在计算所有样本总误差当成一种对数据处理能力的优化，对寻找最优权重过程的PSO算法当成一种对计算的优化。这也是本文分布式下粒子群结合BP神经网络算法的关键点。

综上，分析出了分布式粒子群优化神经网络的关键问题，本文将对两者结合的细节过程做出具体说明。

### 4.3.2 分布式粒子群优化BP神经网络实现步骤

粒子群优化神经网络的分布式实现主要设计思想在于Map阶段对各个个体进行适应度的计算，而这里的适应度就是在神经网络中的总体误差值，Reduce阶段则计算出种群的全局最优值如图4.1所示。在整个算法过程中Map阶段最为耗时，也是整个分布式实现的最主要关键点。以下为分布式粒子群优化神经网络的实现步骤：

1. 随机化M种群个体,并将其写入到HDFS文件中。我们知道，当群体个体数初始化越多时，寻优的概率就越大，并行化处理每个个体更新则是一种加快计算个体速度处理的方式。
2. 将训练集从HDFS文件中读入Map函数中，以格式以键值对形式进行输入，其中key为每行的位移，value为每行样例的值。
3. 对于2中Map函数的输入后，从HDFS文件系统中读取种群相关数据，并以格式图1-1的形式规范化，计算每一行中种群个体的适应值。将粒子ID以及相应的适应值作为Map函数的输出，格式以键值对形式。在Map函数中，读取的所用样例的值，并利用粒子群相关的更新信息求取适应度，在BP神经网络的三层结构向前传输过程中，得到误差值即为此个体的适应度值。适应度是接下来Reduce选择最优值的基础，需要将适应度和神经网络的权值传递给Reduce函数。
4. 所有个体的权值和适应度都会传递给Reduce函数，按照不同个体的适应度大小更新，将其当前最优秀的个体传往下一代。
5. 将的结果存储在HDFS中，作为下一代的迭代的更新值。

在此计算出的将作为下次的迭代值，更新所有的个体，再进行步骤1循环计算，直到满足迭代次数或者满足所有个体的值不在变化时停止迭代。

Map函数的输入值和输出值以对格式进行，输入值中key为粒子的id, value为字符串的形式按以下的格式输入：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ID |  |  |  |  |  |  |  |

表4.2 粒子格式

Table 4.2 particle format

其中id 为每个粒子的标号，表示粒子当前位置，表示粒子的速度，表示粒子的适应度，表示粒子的最佳位置，表示粒子最佳位置的适应值，表示种群全局最佳位置，表示种群全局最佳位置适应值。

1. 分布式PSO Map函数

在Map函数中正如Function(3), 在Map函数中输入键值对，key表示输入的训练样本一行的位移量，value表示样本值信息。这样操作的目的在于能够对训练集多个Map形式读取数据。对于分布在各个机器上的Map函数，我们通过遍历所有粒子，计算出对应粒子的适应度即BP神经网络中的总误差（此时误差还局限在于对应处理的Map数据中）。为了更好的计算更新全局最优解的位置信息，Map输出的key为各个粒子ID值，对应不同的Map函数将在Reduce函数中处理相同的ID。Map函数输出为key，value为粒子的适应度。

1. 分布式PSO reduce1函数

在Reduce函数中正如Function(4)，每个粒子个体都会集中在一个相同的Reduce函数（相同的key值），value值为粒子适应度。此时就可以计算出各粒子的总体误差值即总体适应度。

对于来自于不同Map的粒子，相同ID的粒子适应度相加，从而计算出各个粒子自身的总体适应度。Reduce1函数主要目的在于对各个粒子的自身进行一次更新操作，如果更新后的最优位置的适应度比以前位置的适应度好，则将其相关信息更新到。Reduce输出中key设定的1标号，value为粒子key。再将其输出的作为Map2函数的输入，传入到reduce2函数中。

1. 分布式PSO reduce2函数

在Reduce函数中正如Function(5)，它的目的在于计算出各个粒子的全局最优值，然后粒子通过自身的当前位置，速度以及当前的适应度等信息，更新现在粒子相关信息。在寻找到全局最优适应度过程中，如果一个粒子的比另一个粒子的适应度好，则保持适应度好的，并记录其最优位置。在寻找到全局最优后需要对所有粒子的进行一次有效地更新操作，同时还应对粒子的位置，以及速度信息做一次更新操作。在此过程中我们需要在每次都保存好和。在Reduce函数处理过程中正如Function(5)。

分布式粒子群优化BP神经网络Map函数伪代码：

|  |
| --- |
| Function3: 输入：<key1,value1> ；  输出：<key2,value2>; |
| //计算分布在各个Map粒子的误差    //解析粒子相关信息  //神经网络向前传播过程，计算误差  //输出key , value值 |

分布式粒子群Reduce函数伪代码：

|  |
| --- |
| Function4: 输入：<key2,value2>  输出：<key3,value3>; |
| //计算所有粒子的自身的值      //更新个体自身的值 |

分布式粒子群优化神经网络Reduce函数伪代码

|  |
| --- |
| Function5: 输入：<key3,value3> ；  输出：<key4,value4>; |
| //获取全局最优位置  //将其保存到本地  //保存最佳的位置    //更新所有粒子的全局最优位置  //解析粒子相关信息  //每个粒子从新设置全局最优值  //更新位置和速度    //保存到HDFS以便下次迭代 |

主函数设计方式如下：

本文中的分布式粒子群优化神经网络过程：

在job中运行Map和Reduce函数后将结果传到另外的Reduce函数中并计算得出reduce的输出后，将需要的最佳保存在本地，当训练集训练到一定程度即如果训练样本集网络总误差小于期望的最小精度，或者迭代次数超过设定的迭代次数则从中找出最优的,将其当成BP神经网络训练完成的结构。否则在此进行相应的迭代过程。

### 4.3.3 分布式改进的量子粒子群优化神经网络算法

上一小节中，本文对处理分布式PSO-BP神经网络关键问题进行了分析，并由此有效地进行了对其分布式的实现。在前文中我们还对PSO-BP神经网络进行了量子化的改进，在此将对其进行并行化。

粒子群优化神经网络和量子粒子群优化神经网络的并行过程，在MapReduce中实现Map函数和Reduce函数的计算处理方式相同。不同的点在于量子粒子群中只对粒子的位置进行更新，而粒子群更新过程既包括对粒子的位置也对速度进行更新。鉴于此点的不同，本节在实现分布式量子粒子群优化神经网络过程将继承本文上一节提出的粒子群优化神经网络的实现过程。

具体修改过程如下：

1. 随机化M种群个体,并将其写入到HDFS文件中。这里我们初始化部分只需要对改进的量子粒子群相关变量进行，粒子维度D, ,，以及值。
2. 将训练集从HDFS文件中读入Map函数中，以格式以键值对形式进行输入，其中key为每行的位移，value为每行样例的值。从HDFS中读取粒子群的相关的信息并以格式图1-1的形式规范化。从Function3:Map函数中输出。
3. 对于2中Map函数的输出的结果传递到下一个Reduce1函数中，通过计算每个粒子的总体适应度后，更新粒子的全局适应度 。此时需要把结果传递给Reduce2函数。
4. 在Reduce2 函数中体现出量子粒子群的行为就是对粒子的位置更新不再需要像粒子群对位置和速度更新的方式，而是以量子行为公式进行。Reduce2函数中将会得到每次更新的 。
5. 将的结果存储在HDFS中，作为下一代的迭代的更新值。
6. 在此计算出的将作为下次的迭代值，更新所有的个体，再进行步骤1循环计算，直到满足迭代次数或者满足所有个体的值不再变化时停止迭代。

## 4.4 本章小结

在本章中，首先做了分布式神经网络的必要性的分析，对于这样的一个迫切需求，我们详细介绍了分布式神经网络的具体步骤和实现过程以及其执行流程。对于神经网络进行预测方面收敛慢以及容易陷入局部最优值的这些问题，提出了分布式粒子群优化神经网络的具体步骤和实现过程。不仅于此，为了在预测准确度上的提高，进而提出来分布式的量子粒子群神经网络的优化实现，这不仅减少前面了提及问题出现的概率，更保证了算法在预测方面准确度的提高。

# 第5章 电力负荷预测建模和实验

## 5.1 数据来源及其预测内容概述

本文根据数据类型分别进行短期电力负荷预测和中长期电力负荷预测实验。由于短期电力负荷预测以秒、分钟或者小时为单位容易形成大数据量，因此在处理技术和数据属性整合上不同于中长期电力负荷预测。预测方法上针对其相关特点以及传统预测方法上的缺陷，将对本文所提出的相关算法进行并行化实验。中长期电力负荷将按本文提出的改进算法进行试验。本文所取的数据一部分来自于从国家电网，气象机构等网站爬虫后经过一定处理获取，另一部分数据为项目“电力交易信息采集与数据挖掘技术研究与开发 ”（批准号：2014-0-1-16489）获取。

1. 中长期电力负荷数据类型如表：

中长期电力负荷样本数据本文收集了华中地区几个主要的城市的相关数据集，统计出从2011年至2014年4年的如下表5.1的样例。实验过程中本文将2011年至2013年的36条样例数据作为训练集，同时将2013到2014年的24个样例数据作为测试集。

表5.1 某地区中长期电力负荷样本表

Table 5.1 The Sample list of long-term power load

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 年份 | 价格指数(分/度) | 对外经济(亿) | 工业用电量（Mwh） | 低温高温天数（天） | 钢材消耗（万吨） | 汽车增长量(万辆) | 最大电力负荷(Mwh) |
| 2011.01 105.0 27.80 89.69 20 283.89 17.60 19948  2011.02 105.7 26.89 87.49 11 267.52 18.32 19280  2011.03 105.7 24.45 78.76 1 253.62 20.21 18275  2011.04 105.4 23.41 74.77 1 251.88 13.31 16127  2011.05 105.8 25.97 72.80 5 244.57 14.79 16107  2011.06 106.8 26.66 86.77 10 314.31 18.19 19264  2011.07 106.7 33.94 108.16 20 349.54 24.75 25005  2011.08 106.3 31.11 94.07 22 320.77 22.85 23672  2011.09 106.6 30.78 81.50 19 313.29 20.85 20804  2011.10 106.1 27.15 75.32 8 269.30 15.25 17579  2011.11 104.5 26.14 75.61 4 295.86 16.65 16647  2011.12 104.4 29.93 86.61 18 302.91 21.95 21163  … … … … … … … … | | | | | | | |

1. 短期电力负荷数据类型如表所示：

本文所进行短期电力负荷预测以秒为单位，即15秒对电力负荷进行一次采样。这种采样形式特点在于实时性较高，需要快速对下一阶段可能出现的峰值进行提前预判。实验过程中，本文统计如表5.2所示的大概在26400条的短期电力负荷数据样例。

表5.2 短期电力负荷样例表

Table 5.2 The Sample list of short-term power load

|  |  |
| --- | --- |
| 时间点 负荷量(kwh) | 时间点 负荷量(kwh) |
| 2015/1/1 0:15:00 2902.0989  2015/1/1 0:30:00 2888.1550  2015/1/1 0:45:00 2888.6400  2015/1/1 1:00:00 2849.2449  2015/1/1 1:15:00 2820.3280  2015/1/1 1:30:00 2801.1800  2015/1/1 1:45:00 2811.6850  2015/1/1 2:00:00 2806.2358  2015/1/1 2:15:00 2775.7710  2015/1/1 2:30:00 2751.8070  2015/1/1 2:45:00 2764.7910  2015/1/1 3:00:00 2779.9165 | 2015/1/1 3:15:00 2759.4695  2015/1/1 3:30:00 2754.0244  2015/1/1 3:45:00 2740.9453  2015/1/1 4:00:00 2726.5870  2015/1/1 4:15:00 2769.8910  2015/1/1 4:30:00 2804.5493  2015/1/1 4:45:00 2799.0662  2015/1/1 5:00:00 2788.4060  2015/1/1 5:15:00 2822.6316  2015/1/1 5:30:00 2830.7810  2015/1/1 5:45:00 2840.3142  … … |

## 5.2 评价标准

评价分析误差的标准很多其中，主要本文将以下列的形式进行评价。

1. 平均绝对误差。

其中为平均绝对误差，为实际电力负荷，为预测的电力负荷。这种评价标准从整体上能够对算法好坏进行判定。

1. 均方误差。

其中MSE为均方误差，为实际电力负荷，为预测的电力负荷。这种评价标准可以对结果的好坏进行判定。

1. 精度

其中k为实际电力负荷与预测电力负荷在相对误差范围内的个数。n为所有样例的个数。Precision就是本文进行性能评价算法总体精度。

1. 总体迭代次数

其中为在BP神经网络中能够使平均绝对误差能达到阈值需要经过的迭代次数，k2优化的粒子群算法寻找到使MAE达到最小的最优权重的迭代次数。

1. 加速比

其中代表为算法在非并行化部署运行时间，代表并行化算法部署运行的时间。

## 5.3 电力负荷模型建立

### 5.3.1 输入关联度预处理

对于电力负荷的预测，特别是中长期电力负荷预测所受的影响因素多种多样。因此选择合理的电力负荷影响因素即BP神经网络的输入，将直接影响BP神经网络的预测准确性。本文对影响因素的选取采用灰度关联法。灰度关联法可以对各因素进行评估，留取对目标值相关的影响因素，从而将其作为BP神经网络的输入。具体处理过程如下：

1. 确定分析数列。

首先将电力负荷作为参考数列，可以使用 表示，其中表示电力负荷时段的数值。然后通过选取电价指数，工业用电量，以及天气等因素作为比较数列，比较数列可以使用表示，其中表示某个影响因素某个时段的值，j表示影响因素个数。

1. 变量的无量纲化处理。

由于电力负荷预测系统中的各个影响因素的量纲不同，在计算以及比较时难以得出正确的结论。因此在进行灰度关联分析时，一般采用如下公式进行无量纲化处理。

其中i取值为1到n。

1. 计算关联系数。

对变量进行处理完之后，采用下列公式进行关系系数的计算。

其中表示参考数列与影响因素A之间的关联系数，表示参考数列与比较数列的绝对值，表示参考数列与比较数列的绝对值中最小值，表示参考数列与比较数列的绝对值中最大值，一般取值为0.5。

4．确定影响因素。

计算平均关联系数，从中选取大于阈值的影响因素作为真正的影响因素，本文选取阈值为0.6。从而选出如图的输入变量即影响电力负荷影响因素。

### 5.3.2 数据归一化处理

1.数值型数据处理

对用例数据的归一化处理，目的在于消除变量之间的量纲影响。一般情况下的数据归一化处理采用如下公式：

一般情况下，采用如上公式计算结果会出现某一影响因素值为0的情况，这种情况下并不符合本文对数据的处理。因此归一化的形式按自定义公式进行：

其中，为自定义的阈值。目的在于对为0值进行填充。

2.标量型数据处理

本文针对气候变化对电力负荷影响特点，采用如下方式进行处理：将某地区气候因素散列成三个区间（低温，适度温度（10-26），高温），统计出的温度（低温，高温）天数，将低温天数用表示，高温天数用表示，映射到0-1之间公式如下：

其中n为当月的总天数。本文针对天气进行如上考虑原因在于：“高温”，“低温”都会对电力负荷的变化产生影响，而“适度”对其影响不是很大，使用如上处理正好符合现实要求。

### 5.3.3 参数设置

1. BP神经网络参数设置：

对于BP神经网络中有一条重要的性质，对任何区间上的连续函数，都只需要使用三层结构进行逼近。因而本文同样选用三层网络结构即输入层，输出层和隐含层。BP神经网络的输入和输出的节点数目确定，通常是由输入的属性个数和预测个数确定，在本实验中，我们选用了6个输入属性，目标数据的维度为3。对连续值的预测，一般情况下BP神经网络的输出层节点为1，本文在优化方面做出了改进即将输出层节点个数设置为3，这样做的目的在于增加网络的复杂度，增强预测能力。这样做的可行性：在预测电力负荷时，将电力负荷预测的误差首先进行离散在一个可容忍的误差范围分别设定在三个输出层网络节点上。各层之间的激活函数都选定为Sigmoid函数。综上所述，本实验选用的三层BP神经网络中，输入层节点个数，输出层节点个数分别为6和3。

隐含层节点个数一般选取方式为按照经验或者是已有的实验进行选取。隐含层节点数目个数一般跟输入层，输出层个数有直接关系。隐含层节点数目过多会使学习时间过长，效果不一定好。一般采取的求解隐含层节点个数公式如下：

其中N表示为隐含层个数，M表示输入层个数，Q表示为输出层个数，a取[1,10]之间的常数。本实验在建立隐含层个数时进行了相应的实验。实验结果表明，N取值为10时，此时的BP神经网络的效果达到预期效果。

首先我们通过5-1确定N的范围在[4,13]之间，然后通过BP神经网络实验确定符合的隐含层个数。结果如下表：

表5.3隐含层与误差关系

Table5.3 The relation of Hidden layer and err

|  |
| --- |
| 隐含层个数 4 6 8 10 12 13 |
| 误差值 0.009 0.0043 0.0035 0.0022 0.023 0.0037 |

通过以上对样本在BP神经网络算法中迭代10000次最终能达到的误差值，最后本文选定了10作为隐含层的个数。

1. 粒子群优化参数设置

粒子群中参数有种群规模M,惯性因子，学习因子和 和区间速度，在本论文中为了种群规模M=100；粒子维度值按公式3.1中D计算。速度大小决定了当前粒子搜索过程。如果太慢，粒子则不能再局部点之外进行对全局最优值进行搜索，如果太快，粒子可能在快到达极值点时候丢失。本文设置的。

通常情况下，学习因子和一般取相同的值2。在论文研究表明：和为常数时能求得比较好的解，但是它的值不一定需要取2。在引入收敛因子来保证粒子群的收敛性。实验表明，如果取得恰当的惯性因子，学习因子和，它的收敛的速度将同使用相差不大。因此本文选取的子，。算法终止迭代次数设置的尽可能高，目的是为了实验对算法的性能进行有效地测试。

1. 量子粒子群参数设置

在量子粒子群中参数相对比粒子群的参数更少，只有一个被称为收缩-扩展因子。在论文[8]中实验得出的<1.782时粒子收敛。因此可以对其取值时进行在[0.5,1.782]之间递减进行处理。

1. 改进的量子粒子群算法参数设置

在改进的量子粒子群算法中，我们对量子粒子群中的收敛系数进行了重新定义，如同 3-7式子中。在研究相关文献对收敛性范围基础上引入收敛因子，使其在一定范围内线性递减。同时对根据本文对早熟的定义，对早熟的参数进行相应的设置。

## 5.4 中长期电力负荷相关优化算法实验

接下来本文将会对上述表5-1预处理过后的样本数据分别进行三种算法

的测试，整个测试分析主要包括以下几个方面：

1. 测试PSONN算法，QPSONN算法以及XQPSONN算法在训练集上迭代

次数与评价函数的关系，同时详细分析XQPSO相对其他两种算法优化BP神经网络在迭代次数，以及寻优方面的优势。

1. 测试PSONN，QPSONN以及XQPSONN三种算法预测测试集时真实值

以及预测值之间的拟合效果，同时展示实验过后预测归一化时的误差值。本文重点选取了2014年度的电力负荷实际值与三种算法的预测值进行了相应的拟合效果。

1. 测试PSONN，QPSONN以及XQPSONN三种算法整体的相对预测性能。

本文选取了每次实验结果最好的准确度，平均误差绝对值以及平均方差三个评价指标进行相应的分析。

对于三个BP神经网络的优化算法在预测中长期的电力负荷实验中的寻优迭代次数变化关系，本文将PSONN，QPSONN以及XQPSONN三种算法分别对训练集进行了8000次迭代训练，得出了如图所示对比情况。

图5.1 PSONN,QPSONN,XPSONN迭代次数与误差曲线

Fig 5.1 PSONN,QPSONN,XPSONN Iteration number and error curve

从上图5.1中可以看出，针对训练集训练过程中，三种优化算法都进行了不断寻优过程。其中QPSONN算法以及XQPSONN算法在寻优过程的初期（如上图的迭代20次之内）就出现了适应值急剧下降的情况，这说明其在初期就会就能对BP神经网络模型进行寻优，并且寻优变化量大、效果好。PSONN算法则在此期间适应值仍然保持相对持平，说明该算法在初期寻优并不敏感，这也侧面说明了其在收敛性上没有QPSONN算法以及XQPSONN算法好，反映了本文在加入惯性因子后算法收敛性更强。在算法迭代中期过程XQPSONN算法适应值始终比PSONN算法适应值小，QPSONN算法在迭代过程中一段过程适应值超过了PSONN，这说明了XQPSONN算法在迭代中期也能保持一定的稳定性。算法迭代后期，三种算法都能达到一个稳定值，这说明了三种算法最后都能收敛，但是XQPSONN最后的收敛结果比其他的算法结果更好，说明了将收敛因子进行递减式寻优得出的效果更好。最后如果我们设定一个预测适应度阈值，从上图中可以看出XQPSONN只需要45次迭代，而其他算法则需要90次以上迭代实现。总之，XQPSONN算法不管在寻优迭代次数上，还是在最后的适应值效果上都会优于PSONN和QPSONN两种算法。

下面针对的是三种算法对测试集中电力负荷预测的拟合效果图。本文选取了2014年度的实际值与预测值进行了相应的拟合。在迭代终止条件上本次实验选取的总体样本的误差阈值为0.01得出的效果如图5.2所示。在此基础上论文中还列出了三种算法对2014年度电力负荷预测结果值在归一化情况下和实际值之间的误差情况，如表5.4中所示。

图5.2 2014年度的电力负荷实际值预测值拟合效果图

Fig 5.2 Fitting effect diagram of actual value prediction

表5.4 优化算法归一化时相关预测结果对比

Table 5.4 Comparison of normalized result

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 月份 | 实际值 | PSONN算法 | QPSONN算法 | XQPSONN算法 |
| 预测值 误差 | 预测值 误差 | 预测值 误差 |
| 1 0.695 0.637 0.058 0.632 0.063 0.701 -0.006  2 0.556 0.512 0.044 0.452 0.104 0.511 0.045  3 0.436 0.394 0.042 0.346 0.090 0.409 0.027  4 0.296 0.325 0.029 0.338 -0.042 0.287 0.009  5 0.297 0.324 -0.027 0.342 -0.045 0.278 0.019  6 0.698 0.633 0.065 0.650 0.048 0.684 0.014  7 0.937 0.852 0.085 0.819 0.018 0.894 0.043  8 0.940 0.816 0.124 0.846 0.094 0.880 0.060  9 0.462 0.335 0.127 0.479 -0.017 0.497 -0.035  10 0.177 0.177 0 0.283 -0.106 0.163 0.014  11 0.289 0.186 0.103 0.350 -0.071 0.266 0.023  12 0.620 0.576 0.044 0.584 0.036 0.592 0.028 | | | | |

从图5.2的实验结果可以看出，三种算法在大体的趋势上和2014年度的电力负荷预测的真实值是相适应的。对三种算法分别进行相应的分析来看，PSONN算法在与真实值之间的对比来看效果相对最差，从表中5.4的误差中反映出每个样本预测的值的准确度很少能够达到0.03以下，它的平均误差在0.063上下，这也就说明了其预测的结果并不是每个样例都很准确。从图中5.2，XQPSONN算法预测的结果相对QPSONN算法拟合的实际值更好，基本上是顺应实际结果的趋势，在误差精度阈值在0.03以下的样例个数明显比其他两个的样例多，其平均误差在0.034上下。从中对比我们可以看出XQPSONN在预测结果值中，每个样例的预测相对于其他两个算法效果更佳好，总体上看效果也更加优异。

下面表5.5表示的是三种算法在一定次数的重复实验中所能达到的最优性能情况。本次实验依据平均准确率，平均误差绝对值以及均方误差标准，来评价各个算法性能的好坏。

表5.5 算法性能比较

Table 5.5 Algorithms performance

|  |  |  |  |
| --- | --- | --- | --- |
| 优化算法 | 准确率 | 平均误差绝对值 | 均方误差 |
| PSONN 70% 0.0626 0.03758 | | | |
| QPSONN 74.8% 0.0613 0.00379 | | | |
| XQPSONN 85% 0.0432 0.00321 | | | |

在表5.5实验所示，平均准确率是指在设定了一定阈值（本文选取为0.01）后，每个预测的值与实际值之间的差值，如果小于阈值则就认为其预测的符合期望，从而统计出相应的符合期望的个数，算出平均准确率。均方误差值是指适应度能达到的最优值，它从整体上对反映了所需要预测的所有样例的误差值相对大小。平均误差绝对值也可从整体上反映出算法整体的精度好坏。从上表我们可以看出XQPSONN在准确率上比其他两个算法多出了10个百分点，在均方误差方面较其他二者更加优异。这说明了改进后的算法在整体准确度上确实有一定的提高，同时稳定性比较其他两个算法也更加的优异。总之XQPSONN算法在性能上有一定的提升。

## 5.5 短期电力负荷分布式优化算法实验

为了验证Hadoop环境下并行BP神经网络以及相关的BP神经网络优化算法的有效性，本章将进行相应的对比试验。试验中我们将数据分为两种：一种数据量将近2M，样例条数为26400条，另一种为人工仿真的数据，大概的数据量在700M。在实验过程中我们对单机的BP优化算法和Hadoop环境下BP优化算法在训练时间上进行对比，实验目的在于检测算法是否有较高的扩展性。整个实验包括以下几个部分：

1. 将各个对BP神经网络改进的算法部署了Hadoop平台上，分别对分布式BP神经网络，PSONN，QPSONN，XQPSONN算法在不同数据集的收敛时间进行测试，并详细的解析其出现的原因。
2. 测试分布式BP神经网络，PSONN，QPSONN，XQPSONN算法在不同集群节点上的加速比，展示出改进后的算法在部署分布式扩展性。

### 5.5.1 分布式平台搭建

对于hadoop集群的搭建，我们使用4台机器组成集群来测试地第四章我们提及的相关的BP神经网络分布式优化算法。相关节点配置如表 所示，其配置信息包括软硬件以及各种网络连接的配置，硬件配置信息包括各个节点的CPU，硬盘信息，内存等，软件配置主要包括每个节点所搭配的操作系统以及实验测试所必须的各种软件配置，网络连接配置中设置各个节点的名称，以及各个节点的IP等信息。本论文的相关分布式实验主要以上述集群为实验环境进行，对分布式BP神经网络算法，分布式PSONN算法，分布式QPSONN算法，以及分布式XQPSONN算法在短期电力负荷数据上进行具体的应用实验。实验的主要目的在于对其非并行化时处理数据效率对比，同时对最后的预测结果进行相应的分析说明。各节点的配置信息如下：

表5.6 集群节点配置信息

Table 5.6 The configuration information of nodes in the cluster

|  |
| --- |
| 参数名称 配置情况 |
| CPU Intel I5-4690 3.3GHZ 4Core  网卡 1000Mbps以太网  普通硬盘 1TB 7200 转/分  固态硬盘 三星120G  操作系统 Ubuntu 14.04 LTS  云平台 云平台  JDK openjdk-7-jdk |

表5.7 各集群节点的部署信息

Table 5.7 The deployment information of the cluster

|  |
| --- |
| 主机名 IP Mac地址 |
| Hadoop0 202.118.11.99 4439c49647cf  Hadoop1 202.118.11.106 4439c495ffb6  Hadoop2 202.118.11.32 6c0b8403f4e3  Hadoop3 202.118.11.16 6c0b8403f24c |

### 5.5.2 模型建立与实验分析

对于下一时间段的提前预测，本文采用的方式按照以前电力负荷的时间序列来预测下一段的电力负荷。神经网络中的输入层设置为4，输出层设置为1。即我们可以将前4个15秒的电力负荷作为输入值，对下一个15秒进行电力负荷预测实验。输入形式如下：

从而可以形成如下样例形式，通过归一化处理后对三种优化算法进行训练。

表5.8 短期电力负荷样例表

Table 5.8 The Sample list of short-term power load

|  |
| --- |
| 样例 y |
| 1 2902.0989 2888.1550 2888.6400 2849.2449 2820.3280  2 2820.3280 2801.1800 2811.6850 2806.2358 2775.7710  3 2775.7710 2751.8070 2764.7910 2779.9165 2759.4695  4 2759.4695 2754.0244 2740.9453 2726.5870 2769.8910  5 2769.8910 2804.5493 2799.0662 2788.4060 2822.6316  … … … … … … |

对于参数设定以及数据归一化处理部分，参照5.3节。经过处理后的数据的样例个数取26400即数据量大概为1M，取200000数据量为700M数据分别进行三种优化算法的实验。本次实验结果为每个算法运行结果为10次后取的时间平均值。下表5.9为三种算法在不同数据集的运行收敛时间比较。

表5.9 三种算法收敛时间

Table 5.9 Algorithm convergence time

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 算法名称 | 数据集大小 单机 | | | 1个slave | | 2个slave | | 3个slave | | 4个slave | |
| BP神经网络 | 1M  700M | 1.9s  2052s | 1.8s  2331s | | 1.4s  1200s | | 1.6s  916s | | 1.5s  891s | |
| PSONN | 1M  700M | 1.6s  2032s | 1.4s  2283s | | 1.4s  1181s | | 1.5s  736s | | 1.4s  742s | |
| QPSONN | 1M  700M | 1.3s  2023 | 1.4s  2223s | | 1.3s  1136s | | 1.4s  707s | | 1.3s  709s | |
| XQPSONN | 1M  700M | 1.2s  2021s | 1.3s  2245s | | 1.4s  1135s | | 1.3s  690s | | 1.4s  688s | |

为了测试改进算法在分布式集群中的表现，分别在单机和Slave节点上部署了算法，并统计了各个算法达到收敛时的时间如表5.9所示。在实验过程中分别取了1M真实数据集进行了测试，另外对700M的人工仿真数据集进行了测试。从表中可以看出1M的数据集对于各种算法在不同Slave处理时间基本上是相等的，但是在各种算法中XQPSONN收敛的时间相对其他算法，不管在单机还是集群中都有一定的提升，这一点可以验证我们提出的线性递减的惯性因子对算法快速收敛有一定的作用。从各个算法自身看，其无论在单机还是集群收敛时间都大致一样，这点可以理解为对于小数据量集群的收敛速度跟单机不会有太大的不同。对于700M的数据集各个分布式算法收敛时间如图5.3所示。

图5.3 BP, PSONN, QPSONN, XQPSONN算法在集群收敛时间

Fig. 5.3 Algorithms Convergence time in cluster

图5.4 算法加速比

Fig 5.4 speedup ratio

图5.4展示了700M数据集时各个算法在不同数量节点的收敛加速比。从中我们可以看见几种分布式算法都有较好的加速过程，尤其分布式XQPSONN算法在3个slave节点时加速比接近3。随着Slave的增加，加速比增长率会出现明显的下降趋势，出现一个平稳曲线，这主要是因为随着集群规模的增大，各节点之间进行通信量也会增大，同样对硬盘的读写数据量也会频繁，从而影响了整体集群的性能。

## 5.6本章小结

本章节中，首先对本文实验的数据集进行了介绍，同时还对数据的相关处理过程进行了介绍。然后对实验的各个部分评价标准进行了定义。最后针对电力负荷的特点，分别对中长期电力负荷预测和短期电力负荷预测运用了本文提出的相关算法进行测试。

在中长期的电力负荷预测中，本文主要从算法收敛性，寻优的迭代次数，以及精度等方面进行了实验。实验过程中分别对PSONN，QPSONN以及XQPSONN算法在处理电力负荷预测性能进行了详细的比对。

在短期电力负荷预测中，本文分别实现了各个改进算法的并行化，并分别将BP神经网络，PSONN,QPSONN以及XQPSONN算法的部署到Hadoop平台中，对其进行了收敛时间的测试，同时计算出各个算法的加速比，验证了各个算法的扩展性。

# 第6章 总结与展望

## 6.1 总结

论文首先介绍了电力负荷预测的相关背景，以及研究目的意义，接下来介绍了数据挖掘中对电力负荷预测的相关技术工作。通过了解国内外对电力负荷方面的研究，确定本文以BP神经网络为研究基础，改进BP神经网络在寻优上的不足从而进行电力负荷预测。本文在主体部分首先详细介绍了BP神经网络算法模型原理等，进一步介绍了对BP神经网络优化的技术粒子群优化神经网络，同时为对算法的进一步优化引入量子技术在粒子群方面的技术思想。然后本文通过详细分析上述相关技术的局限性，总结上述各种算法在寻优方面的优点，基于粒子群以及量子思想，提出了基于量子粒子群优化神经网络的改进算法。在此过程中本文又对粒子群与神经网络的结合进行了详细描述，从而为后续改进算法提供借鉴意义。本文还考虑到面对当今社会海量数据，分布式处理的热点问题，论文介绍了目前流行的分布式处理框架MapReduce,介绍了整个分布式框架的运行机制和原理。为解决传统算法对电力负荷大数据的不足之处，本文提出了分布式算法处理海量数据的措施，为此本文提出来基于MapReduce分布式优化算法。本文主要做了以下几项工作：

1. 根据BP神经网络在电力负荷上面的不足，应用PSO优化BP神经网络预测电力负荷，同时针对传统预测对BP神经网络输出层单一，考虑误差容忍度，将其设定成三个节点结果，增强了网络预测效果。
2. 为解决粒子群优化过程中的聚集效应，以及收敛性问题。根据粒子群以及量子特点结合BP神经网络，提出来改进的量子粒子群优化神经网络算法，其大大的减少了迭代步数，以及提高了寻优的结果。
3. 针对电力负荷的海量数据特性，分析了实现分布式神经网络优化算法的必要性，从而提出了改进的相关分布式算法预测电力负荷。为此本文设计出了实现分布式的BP神经网络MapReduce函数，同时实现了相关改进算法的分布式设计，实验表明他们对电力大数据进行了很好的处理。
4. 针对电力负荷预测的特点，本文采用了灰度关联法对影响因素进行了提取，确定了关联度高的因素，同时分别对中长期电力负荷预测和短期电力负荷预测进行了实验工作。在对中长期电力负荷预测时，本文应用了多个外部因素如电价，工业用电量，钢铁量等进行了相关实验，并取得了良好的拟合效果。短期电力负荷预测过程中，本文为说明算法的扩展性，对分布式算法进行了加速比实验。

## 6.2 展望

本文提出以外部影响因素如电价，钢铁量，国家宏观经济状况等对电力负荷的预测，结合粒子群优化和量子特性，提出了改进的量子粒子群优化神经网络，提高了收敛速度以及寻优精度。为处理电力大数据，本文将相关算法部署在MapReduce分布式平台，提高了处理数据的效率。通过以上的研究，发现了上述工作仍然存在许多待解决的问题：

1. 对中长期电力负荷影响的外部因素应远远不止上文考虑的，本文只是对能取得相关因素进行了实验，，如果能对影响因素细致的划分运用上述过程效果应该更好。对短期电力负荷的由于其数据采集的时间段，变化大，因此本文未对其进行相应的因素影响考虑，只是单纯的进行了外推的方法将其作为BP神经网络的输入，有一定的缺陷。
2. 对于上述进行了各种算法预测性试验过程中，由于各个算法本身的参数比较多，很难去一一的确定最好的参数。本文更多的参照了相关文献出现比较好结果时的参数值进行了实验，因此一定有更好的参数调优对实验结果产生影响。为此仍需要接下来花更多时间进行参数调优。
3. 基于MapReduce分布式平台上实现的各种优化算法，借助了多台机器对数据处理效率上优化，但是并没有真正的去改进算法本身对数据处理优化。由于时间和精力的原因，我们在本文并没又给出一个很好处理方案，在今后的工作中，将继续探讨。

# 参考文献

[1] 刘学琴, 吴耀华, 崔宝华. 基于扩展粗糙集的短期电力负荷预测模型. 电力系统保护与控制, 2010, 38(5):25-28.

[2] 乔维德. 基于改进PSO-BP神经网络的短期电力负荷预测研究. 电力系统保护与控制, 2007, 35(17):17-21.

[3] 莫维仁, 孙宏斌, 张伯明. 面向电力市场的短期负荷预测系统设计及实现[J].

[4] Heiko Hahn, silja Meyer-Nieberg, Stefan Pickl. Electric load forecasting methods: Tools for decision making. European Journal of Operational Research 199(2009)902-907.

[5] 牛东晓, 陈志业. 组合灰色神经网络模型及其季节性负荷预测[J]. 华北电力大学学报自然科学版, 2000, 27(4):1-6.

[6] Diego J.Pedregal, Juan R.Trapero. Mid-term hourly electricity forecasting based on a multi-rate approach Energy. Conversion and Management 51(2010)105-111

[7] Jinbo Shang, Yu Zheng, WenZhu Tong, Eric Chang, Yong Yu. Inferring Gas Consumption and Pollution Emissions of Vehicles throughout a City.  Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014.

[8] Xinwei Wang, Yanpeng Cai, Jiajun Chen, Chao Dai. A gey\_forecasting interval-parameter mixed-integer programming approach for integrated electric-environmental. management-A case study of Beijing Energy63(2013)334-334.

[9] Min Jin, Xiang Zhou, Zhi M.Zhang, Manos M.Tentzeris. Short-term power load forecasting using grey correlation contest modeling. Expert Systems with Application 39(2012)773-779.

[10] White, Halbert. "Economic prediction using neural networks: the case of IBM daily stock returns." IEEE International Conference on Neural Networks 1988:451-458 vol.2.

[11] 高隽. 人工神经网络原理及仿真实例[M]. 机械工业出版社, 2003.

[12] 韩力群. 人工神经网络理论、设计及应用[M]. 化学工业出版社, 2007.

[13] 电力系统自动化, 2001, 25(23):41-44.

[14] 李正学. 神经网络理论及其在股市短期预测中的应用[D]. 吉林大学, 2001.

[15] 贾文臣, 叶世伟, 陈国梁,等. 前馈神经网络凸优化算法及其仿真应用[J]. 计算机仿真, 2004, 21(9):113-116.

[16] 崔和瑞, 彭旭. 基于ARIMAX模型的夏季短期电力负荷预测. 电力系统保护与控制, 2015, (4):108-114.

[17] 毕洪波, 张玉波. 混沌RBF神经网络在电力负荷预测中的应用. 科学技术与工程, 2009, 9(24):7489-7492.

[18] 徐明义. 混沌神经网络在短期电力负荷预测中的应用. 信息技术, 2010, (11):49-51.

[19] 孔俊, 吴微. Online Gradient Methods with a Punishing Term for Neural Networks[J]. Communications in Mathematical Research, 2001, 17(3):371-378.

[20] Wei Chiang Hong. Application of chaotic ant swarm optimization in electric load forecasting Energy Policy 38(2010)5830-5839

[21] Clerc, M. "The swarm and the queen: towards a deterministic and adaptive particle swarm optimization." Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on IEEE, 1999.

[22] 王丽, 王晓凯. 一种非线性改变惯性权重的粒子群算法[J]. 计算机工程与应用, 2007, 43(4):47-48.

[23] Jiang M, Gielen G, Bo Z, et al. Fast Learning Algorithms for Feedforward Neural Networks[J]. Applied Intelligence, 2003, 18(1):37-54.

[24] Plagianakos V P, Magoulas G D, Vrahatis M N. Deterministic nonmonotone strategies for effective training of multilayer perceptrons.[J]. IEEE Transactions on Neural Networks, 2002, 13(6):1268-84.

[25] 侯云鹤, 鲁丽娟, 熊信艮,等. 改进粒子群算法及其在电力系统经济负荷分配中的应用[J]. 中国电机工程学报, 2004, 24(7):95-100.

[26] 王冰山, 周步祥, 肖贤,等. 基于量子行为粒子群优化算法-混沌神经网络的电力系统负荷预测. 电工电能新技术, 2014, 33(6):7-12.

[27] 孙俊. 量子行为粒子群优化算法研究[D]. 江南大学, 2009.

[28] 王莹莹. 量子粒子群算法及其在贝叶斯网络结构学习的应用研究[D]. 南京邮电大学, 2014.

[29] 徐炜, 肖智, 杨道理. 量子算法在大数据挖掘中的应用前景浅析.2013中国信息经济学会学术年会暨博士生论坛论文集. 2013.

[30] Peng T M, Hubele N F, Karady G G. Advancement in the application of neural networks for short-term load forecasting[J]. IEEE Transactions on Power Systems, 1992, 7(1):250-257.

[31] Coelho L D S. A quantum particle swarm optimizer with chaotic mutation operator[J]. Chaos Solitons & Fractals, 2008, 37(5):1409-1418.

[32] Coelho L S. Novel Gaussian quantum-behaved particle swarm optimiser applied to electromagnetic design[J]. Science Measurement & Technology Iet, 2007, 1(5):290-294.

[33] Li, Shouyi, et al. "A New QPSO Based BP Neural Network for Face Detection." Fuzzy Information and Engineering, Proceedings of the Second International Conference of Fuzzy Information and Engineering, ICFIE 2007, May 13-16, 2007, Guangzhou, China 2007:355-363.

[34] Mikki S M, Kishk A A. Quantum Particle Swarm Optimization for Electromagnetics[J]. IEEE Transactions on Antennas & Propagation, 2006, 54(10):2764-2775.

[35] Fabio Vesperini, et al. "A neural network based algorithm for speaker localization in a multi-room environment." 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP) 2016.

[36] Kedlaya K S. Quantum computation of zeta functions of curves[J]. computational complexity, 2006, 15(1):1-19.

[37] Karafyllidis I G. Quantum Computer Simulator Based on the Circuit Model of Quantum Computation[J]. Circuits & Systems I Regular Papers IEEE Transactions on, 2005, 52(8):1590-1596.

[38] Lin Y A. Map-Reduce for Machine Learning on Multicore[J]. Advances in Neural Information Processing Systems, 2006, 19:281-288.

[39] Marin Matijas, Johan A.K.Suykens, Slavko Krajcar. Load forecasting using a multivariate meta-learning system. Expert System with Application 40(2013)4427-4437

[40] Liu, Zhiqiang, H. Y. Li, and G. Miao. "MapReduce-based Backpropagation Neural Network over large scale mobile data. "International Conference on Natural Computation, Icnc 2010, Yantai, Shandong, China, 10-12 August 2010:1726-1730.

[41] Gençay R. Non-linear prediction of security returns with moving average rules[J]. Journal of Forecasting, 1996, 15(3):165-174.

[42] Chu C T, Sang K K, Lin Y A, et al. Map-Reduce for Machine Learning on Multicore.[J]. Advances in Neural Information Processing Systems, 2006, 19:281-288.

[43] Graves, Alex, et al. "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. International Conference on Machine Learning ACM, 2006:369-376.

# 致 谢

时光荏苒，两年半的研究生终点如约而至。回顾这些时光自己所获得的收获还是蛮多，在学业中掌握了许多不曾了解的知识，更重要的是懂得了些许学术研究方法；在生活中结交了一批能吐露心声的好朋友，是他们教会了我怎么面对逆境中种种不顺，学会应当处以什么样的心态。现在快要离开美丽的校园，步入人生的另一段旅程，正当此时聚集的更多的是一种对人和事的种种不舍。

首先我要感谢我的导师焦明海老师，是他在这两年半的时间无论在学业和生活中都给予我无微不至的教诲和关怀。从本论文刚开始确定方向，到论证题目合适与否，最后内容选取都给予我极大的帮助。在此期间，焦老师不论在学术上教导我刻苦专研，生活中更是叮嘱我要勇于面对困难。焦老师的渊博知识以及乐观的为人都会使我更加的尊敬，也必然会对我以后工作有着很大的影响。

感谢张岩峰老师，张老师在学术上给我带来了不一样的视野，让我明白学无止境的道理。在张老师的每周例会上，让我了解了更多更新的国内外研究成果，并且在我有疑惑的时候能带我一起想问题。张老师谦逊的为人以及严谨的治学态度，都给了我很大的触动。

感谢实验室的所有兄弟姐妹，在这两年半的日子里，无论快乐还是烦恼，都是他们陪伴我，让我感受大家庭的温暖。

最后我要感谢我的父母， 正是他们的鼓励给予我前进的动力。

# 攻读硕士学位期间的论文和项目情况

参与项目

1. 2014.12-2017.12，教育部留学回国人员科研启动基金项目：基于服务的云数据资源优化关键理论与算法研究(项目编号：49-1)。

2. 2015.01-2016.12，企事业单位委托科技项目：电力交易信息采集与数据挖掘技术研究与开发 ，(项目编号：2014-0-1-16489)。

撰写论文

1. Minghai Jiao, Zhubing Wang, Xu Feng. Quantum Behaved Particle Swarm Optimization Neural Network for Power Load Forecasting, Chinese Control and Decision Conference (CCDC) 2017, May.