

基于内存计算的电力负荷预测^①

杨栋枢¹, 张明明², 石磊¹

¹(国网信息通信产业集团 安徽继远软件有限公司, 合肥 230088)

²(国网江苏省电力公司 信息通信分公司, 南京 210000)

摘要: 内存计算技术的提出和发展, 是基于实际情况的需求. 对诸多行业来说, 其在数据处理方面存在各种各样的问题及困难, 诸如数据处理量极大、数据处理效率偏低、处理速度慢等, 电力行业的负荷预测也遇到阻碍, 要对大批量的数据实时分析做出预测成为一大难题, 本文就以基于内存计算, 结合 BP 神经网络预测模型, 研究负荷预测中的问题, 实验证明比传统方法有质的提升.

关键词: 内存计算; 电力负荷; 预测; BP 神经网络; 大数据

Power Load Forecasting Based on Memory Computing

YANG Dong-Shu¹, ZHANG Ming-Ming², SHI Lei¹

¹(State Grid Information & Telecommunication Industry Company Limited, Anhui Jiyuan Software Company Limited, Hefei 230088, China)

²(State Grid Jiangsu electric power company, Telecommunication branch, Nanjing 210000, China)

Abstract: Proposing memory computing and the development of technology, is based on the actual situation needs. For industries, its data processing is facing all kinds of problems and difficulties, such as large amount of data processing, low efficiency of data processing and slow processing speed, load forecasting of electric power industry also meet obstacles, to large quantities of data real-time analysis make predictions become a big problem, in this paper, the calculation based on memory, combined with bp neural network prediction model, make the research problems on load forecasting, compared with traditional methods, experimental proof has qualitative improvement.

Key words: memory computing; power load; forecasting; BP neural network; big data

现如今, 能源问题日趋严峻, 越来越多的人提倡低碳生活, 而电能人们在生产生活中的地位也愈发重要, 方方面面都离不开电, 是目前能源结构中重要的组成部分之一. 所以电能的使用, 如何使用才能效益最大化, 怎样使用才不会造成过多的浪费, 都是摆在人们面前的诸多问题. 其中, 电力负荷预测也是其中关键的一环, 是电力系统运作中的重要工作之一, 直接关系到电能利用最优化, 效益最大化的问题. 电力负荷预测水平的高低是判断一个电力公司能否实现电力管理现代化重要指标^[1]. 提高电力负荷预测水平, 有助于提高用电管理水平, 有助于增加电力带来的社会效益.

但是在预测电力负荷时, 由于其受到很多因素的影响, 以在风电预测实验的研究为例, 作为世界风电大国, 我国在风电实现跨越式发展的过程中遇到了一系列问题, 主要体现在风电并网消纳受阻, 导致了大量的弃风现象, 以“三北”地区最为突出. 究其原因主要是风电的波动性, 不确定性, 而提高风电功率预测的精度对于风电并网系统中的调度极为有利, 可以促进我国风电消纳的整体水平. 然而, 全球风电功率预测精度普遍不高, 即使是西班牙、丹麦等以风电占主导地位的电力系统中预测精度(误差 15%以内)仍难以达到要求, 寻求高精度的风电功率预测方法是各国亟待解决的问题. 算法精度的提高必然以牺牲实时性

① 收稿时间:2015-11-05;收到修改稿时间:2016-01-27 [doi: 10.15888/j.cnki.csa.005290]

为代价,寻求一种合适的算法实现精度和复杂度之间的折中显得尤为重要。BP神经网络很好的适应了风电预测中的复杂性和不确定性,这一类型的预测具有非线性性和很多因子难以确定,数据量大且十分复杂,一般的负荷数据预测都是需要庞大的数据基础来分析训练神经网络。所以在计算引擎上也提出了更高的要求,在面对海量的业务数据,传统的计算工具显得有点捉襟见肘了。不得不寻求一种高效的计算方法,内存计算逐渐被人们发现与研究。随着内存计算研究的不断深入,将其运用到实际生产中也成为可能。内存计算凭借其运算速度足够快,可以实时预测数据。所以选择预测模型的时候使用基于内存计算的BP神经网络预测算法,以期在获得实时的预测结果还能保证较高的预测精度。

1 内存计算

1.1 内存计算的概述

内存计算是指CPU直接从内存,而不是硬盘上读取数据,并进行计算、分析,内存I/O的读写开销与传统的磁盘I/O读写开销相比,很大程度上得到优化。内存计算非常适合处理海量的数据,以及需要实时获得结果的数据。

随着互联网的不断发展,用户的应用环境越来越复杂,数据变得越来越庞大,动辄百G甚至上T的数据。传统应用设备已经逐渐被生产环境淘汰。内存计算能够使得数据计算的速度呈几何级的增长,带来了充满想象的应用,给用户更过更好的选择。接触过内存计算应用的人普遍表示,借助内存计算,很多以前不可能实现的应用现在都可以实现了。有人认为,内存计算将催生颠覆性的商业模式。

1.2 内存计算的原理及特点

内存计算技术主要是在硬软件系统协作的条件下,将数据库的相关计算转移到内存中进行。也就是说,数据从以往的磁盘迁移到了内存中。在内存中进行数据处理计算,可以有效突破I/O框架的限制,摆脱以往磁盘I/O时空开销,在此基础上还可以进一步采取并行处理技术,大大提高数据处理计算的效率。

内存计算技术的特点如下:首先,在内存层面进行数据读取和处理工作。由于将传统的数据库转移到内存中,摆脱了磁盘I/O的局限性。解决了限制计算效率的一大难题之后,数据处理方式的选择也有了更

优化的选择,从而使数据处理更加开放更加高效。其次,在数据存储层面能够同时并行列储存和行储存,行式存储的数据库在事务处理上十分出色,而列式存储的数据库在数据读取和分析方面更具优势。再者,在处理机制层面也是表现出众,采用并行处理机制。内存计算技术可以支持更多的服务器和处理器,能够分拆数据模型进行分步并行处理,还可以将分拆后的数据模块配置到不同的服务器进行分别处理^[2]。对数据高效压缩优化内存结构。对于电力行业来说,数据库一般都是十分庞大的,只有将这些数据进行高效压缩,优化内存结构,才能提升内存的运转效率,进而提升数据处理效率^[3]。

1.3 内存计算引擎

既然用到内存计算,那么就需要一个强大的内存计算引擎,HANA(High-Performance Analytic Appliance)高性能分析设备。HANA是一套具备软件和配套硬件设备的结合体,提供很多性能卓越的服务,比如说它的数据查询功能,用户可以直接对大量实时业务数据进行查询和分析,而不需要对业务数据进行建模、聚合,并且还能保证其高效性。

在软件方面,HANA的基于内存的数据库(SAP In-Memory Database,IMDB),包括数据库服务器(In-Memory Database Server)、建模工具(Studio)和客户端工具(ODBO、JDBC、ODBC、SQLDBC等)。HANA的计算引擎(Computing Engine)是内存计算的核心所在,负责解析并处理对大量数据的各类CRUDQ操作,支持SQL和MDX语句、SAP和non-SAP数据。这一计算引擎能够快速处理用户复杂的查询请求,快速返回查询结果,相较于传统的查询模式的效率是质的飞跃;从硬件方面来看,SAP和多个硬件厂商合作生产支持HANA的高性能服务器,包括Dell R910、Fujitsu、HP DL580等公司的合作。这些性能卓越的硬件厂商为其提供坚实的硬件支持,硬件配置完全能够适应目前大数据应用场景的需求,为海量数据分析奠定基础。

除了上述的内存计算引擎,HANA还提供工作台,便于引擎的管理和建模,同时还有展示层的功能,将最终结果展示给用户。当然最终要的还是HANA的计算速率很快。它的快在于用大内存提供内存数据库,并在内存数据库里采用列式存储从而可以将更多的数据装进内存(列式存储更适合数据压缩)^[4]。然后使用针对内存的计算引擎来对数据进行分析处理,快也是理

所应当的事。

2 BP神经网络

2.1 BP神经网络的概述

1986年,由两名科学家提出了多层网络的误差反向传播算法理论,即BP算法。这种算法很好的解决困扰科学家们很多年的多层网络中隐含单元的连接权值的难题,这也正式预示着BP神经网络的诞生,为学者们打开了另一扇研究大门。由正向和反向传播两部分构成BP网络的学习过程。通过这两个过程反复运用的,可以逐渐的减少误差,当算法进行到一定的程度之后,误差必然会达到要求的范围之内,通过循环往复误差减小的过程也是一个很耗时的过程,这个耗时是相对来说,比起其他的一些预测模型来说,神经网络的强大的计算能力能够比它们更迅速的计算出最优解。当数据量和实际问题复杂到一定程度后,有时候一个复杂的大数据问题这将是一个漫长的等待,所以精确度和实效性一直是个矛盾的问题。

兼顾时效性和精确度是研究的重中之重,选择神经网络算法就意味着需要付出极大时空开销才能获得所需要的准确的预测结果。或许可以换一种计算模式,来提高其计算效率。这样既可以充分发挥神经网络的高精确度的优势,又能凭借高效的计算模式极大降低计算所需得时间,从而达到时效性和精确度兼顾的目的。

2.2 BP神经网络的结构及特点

BP神经网络算法的流程图如图1。

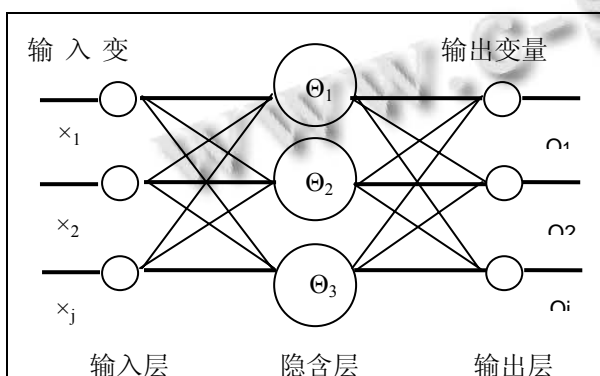


图1 BP神经网络的算法流程图

BP神经网络,能力突出,BP神经网络最主要的优点是具有极强的非线性映射能力。BP神经网络逼近能

力出色,理论上当层数达到三层以上之后可以逼近任何非线性函数,而且拟合出来的函数的精度也是任意可定,也就是说可以人为控制。并且在面对外界刺激或者有输入信号的时候,BP神经网络在处理这些事件时会运用自身的联想记忆的能力^[5]。BP神经网络是采用预先存储信息和学习机制进行自适应训练的。可以从缺失不全的信息和诸多干扰中恢复原始的完整信息。BP神经网络识别与分类能力有不容小觑,尤其在非线性的分类方面,解决了神经网络发展史上困扰人们已久的难题^[6]。另外,BP神经网络其他方面的能力也是十分突出,比如说它的优化计算能力,理论上可以将误差控制在任意精度范围内。

3 内存计算结合BP网络预测设计

3.1 电力负荷预测特点分析

电力负荷预测再实际生产生活中是一项十分复杂工作,充满诸多可变因素,想要预测很难,它是利用过去的的数据来预测未来的用电情况,存在着非常大不确定性,条件及时间性^[7]。比如天气、温度、节气等等。

电力负荷预测具有不确定性和时间性等诸多特点。负荷预测的不确定性表现在它受到多种多样的复杂的因素所影响,而且这些因素还是不断发展变化的,所以预测结果并不唯一。但在负荷预测系统中,不同的场景对不同的负荷预测都有一定的精度要求。长期的负荷允许误差在15%以内,中期的负荷预测的允许误差为5%,而短期的日负荷预测误差一般不能超过3%,日负荷预测各个时间段均方差不超过4%^[8]。

负荷预测的时间性体现在电力负荷预测是在一定的条件下进行的,预测不是空穴来风,是需要严谨的理论支持,加上完整合理的数据分析得出结论,也是属于科学的领域。一般来说,电力企业要根据电力预测实际需要确定电力预测的具体时间,长期电力负荷预测时间一般为2-5年,中期电力负荷预测一般为半年到2年,短期一般为一个月或者一个季度,超短期为一天或者一周,周期越短,要求数据实时分析的性能也越高。相反周期越长,积攒的数据量也越大,对数据分析工具的处理能力要求也越高。综合各方面的因素,内训计算模式配合神经网络预测模型是完全可以胜任大多数的业务需求。

3.2 电力负荷预测数据组成及预处理

实验我对同一组数据进行预测仿真。数据组成:在

一天之中各个时间段的电力负荷的值, 24 小时这里选取 12 组值再加上其他的一些影响因素, 组成本次实验的样本. 实验数据选取了某市 2001~2010 年十年的负荷预测数据, 将采集到的数据先进行归一化处理, 公式如下:

$$y = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (1)$$

截取部分经过归一化处理过后的数据如表 1.

表 1 归一化处理过后的数据

1	0.2452	0.2217	0.2016	0.2115	0.2113
2	0.1446	0.1581	0.1105	0.1201	0.1212
3	0.1314	0.1408	0.1243	0.1312	0.1305
4	0.2246	0.2304	0.1978	0.2019	0.1819
5	0.5532	0.5134	0.5021	0.5532	0.4952
6	0.6642	0.5312	0.5232	0.5736	0.5312
7	0.7015	0.6819	0.6819	0.7029	0.6886
8	0.6981	0.7125	0.6952	0.7032	0.6898
9	0.6821	0.7265	0.7015	0.7189	0.6999
10	0.6945	0.6847	0.6825	0.7019	0.7323
11	0.7549	0.7826	0.7825	0.7965	0.7721
12	0.8215	0.8325	0.7895	0.8025	0.7956
13	0.2415	0.2385	0.2352	0.2542	0.2234
14	0.3027	0.3125	0.2506	0.3125	0.2977
15	0	1	0.5	0	1

(注: 1~12 是一天 24 小时每隔两小时取一个参数, 13 是温度数据, 14 是湿度数据, 15 是天气数据.)

3.3 BP 神经网络预测模型的设计

(1) 网络节点数和层数的选择

隐含层节点数的选择是由实际应用时的具体问题决定的, 并非随便设置的, 需要具体问题具体分析. 隐含层神经元个数选取时也有一定的要求, 需要遵循这些限制: 网络隐含层神经元个数不能太少. 如果太少的话, 无法建立复杂的映射关系, 使网络训练不理想, 直接导致很多后果, 比如对以前没有的样本的识别性能不高, 容错性能有限; 如果神经元个数过多, 会明显增加网络学习时间, 间接导致最终所得的误差有可能不是最小. 综合上述, BP 网络肯定存在最佳的节点数. 对此, 鉴于尚无理论指导, 也采用实验方法进行确定. 实验中可以采用以下公式^[9]来选取节点:

$$h = \sqrt{N + O} + a \quad (2)$$

$$h = \log_2 N \quad (3)$$

隐含层数的选择, 增加隐含层的层数, 会增加网

络的培训时间, 但是相应的会提高网络的泛化能力^[10], 提高网络的精确度, 所以实际运用时要考虑好精度和时效性的均衡.

本次实验输入数据 15 组, 输出节点需要 12 个, 根据上述公式方法, 隐含层节点可选择 31 个. 其中网络的隐含层的选取的传递函数为 S 型正切函数, 输出层选取的传递函数为选择 S 型对数函数, 由此构建 BP 网络, 结构框图如图 2.

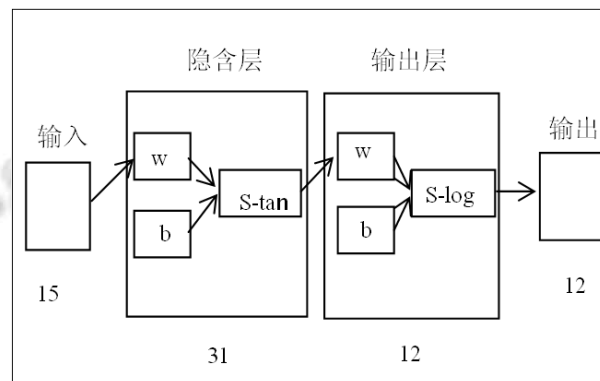


图 2 神经网络模型结构

(2) 训练函数的选择

标准的 BP 神经网络的算法收敛速度慢, 学习速率不易确定. 目前也有很多优化算法来解决此问题, 本次研究采用其中收敛速度最快, 鲁棒性最好的 LM 算法来进行优化^[11]. LM 算法能够根据迭代的结果动态调整阻尼因子来调整收敛方向, 使误差值下降. 但是由于算法对内存的要求很高, 在处理海量数据时效果很差. 但是随着内存计算的发展, 可以解决对内存要求高的问题^[12].

3.4 基于内存计算引擎建立模型设计

在内存计算引擎 HANA 中构建 BP 网络预测模型, 来进行预测分析. 其中数据是存放在 HANA 自身的数据库引擎中的, 这两个数据库都是内存数据库, 所以数据读取存放都是在内存中的. 理论上比传统从硬盘读取数据快 100 万倍^[13], 虽然实际运用中会有诸多限制致使没办法达到这种速度, 但是相较于传统的数据库读取有了革命性的进步.

在 HANA 中建立预测模型, 关键的一个工具是 HANA 提供的预测分析库, 简称 PAL^[14]. 它是一个函数库, 主要用于数据预测和分析, 提供了很多数据挖掘的算法实现. 实验的开发环境是基于 R 语言^[15]的,

在 HANA 的内存数据库中使用 R 语言作为存储过程的开发语言,调用 R 中的统计函数,用于统计分析、绘图和数据挖掘。

4 实验方法及结果分析

在 HANA 中采用基于内存数据库的网络预测模型和基于普通数据库的网络预测模型进行对比试验,先用 2001 年的数据分析预测,然后将 2001 年数据和次年数据合并在做分析预测实验,依次类推,数据量逐渐递增,看对模型的预测性能影响,观看并记录预测效果。

实验通过基于内存数据库的 BP 网络预测和基于普通的数据库的预测进行对比验证。通过对比实验发现,对大批量的数据分别使用两种预测模型来进行预测对比试验时,前者速度有显著提升,在这十年的样本数据分析下,预测时间上减少了 48.6%,并且随着数据量的不断提高,其也测所用的时间增长速率也是相对较慢,也就是说对于海量数据的处理内存计算更具优势。

实验结果如图 3。

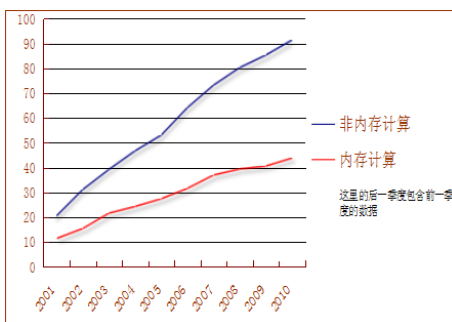


图 3 实验结果比较

5 总结与展望

本文主要研究了基于内存计算的电力负荷预测,利用内存计算对海量数据处理的高效性,适合实时的大数据分析的特点,结合 BP 神经网络对复杂情况的和非线性数据模型的预测能力,来解决电力负荷预测中复杂的业务场景,实验表明方案可行且十分理想。除了满足准确度的要求之外,很能满足海量数据实时分析的需求,提高电力负荷预测水平,有助于提高用电管理水平。电力负荷预测作为电力系统重要的工作之一,同时也是衡量电力系统是否具备先进管理水平

的一项重要指标。其水平的高低意味着是否能让电力能源给社会创造更多得的经济效益。随着研究的不断深入,基于内存计算的神经网络预测在实际生产生活中得到更广泛的运用。

参考文献

- 1 牛东晓,曹树华,赵磊.电力负荷预测技术及其应用.北京:中国电力出版社,1998.20-93.
- 2 罗铮.浅析大数据时代的数据挖掘与数据资产运营.电子技术与软件工程,2014,1:212-213.
- 3 Li Q, Bai XM, Zhang L, et al. Data asset management and data-based operation of power suppliers. East China Electric Power, 2013(1): 324-327.
- 4 李超,张明博,邢春晓,胡劲松:列存储数据库关键技术综述,计算机科学,2010,37(12).
- 5 闻新,周露,李翔,张宝伟.MATLAB 神经网络仿真与应用.北京:科学出版社,2003.
- 6 郝中华.BP 神经网络的非线性思想.洛阳师范学院报,2008,27(4):51-55.
- 7 王钰,郭其一,李维刚.基于改进 BP 神经网络的预测模型及其应用.上海:计算机测量与控制,2005(1):39-42.
- 8 方方.基于 BP 神经网络的电力系统负荷预测.http://www.docin.com/p-815372098.html.[2011-3].
- 9 王立柱,赵大宇.BP 神经网络的改进及应用.沈阳师范大学学报(自然科学版),2007(1):61-64.
- 10 Yan H, Yu SL, Li YL, et al. A design method of the parameters of fractional order PID controller polesorders searching method. Information and Control, 2007, 36 (4) : 445-450.
- 11 武美先,张学良,温淑花,李海楠.BP 神经网络的改进.太原科技大学学报,2005(2):120-125.
- 12 Bayer R, Mc Crwright E. Organization and maintenance of large ordered Indexes. Software Pioneers: Contributions to Software Engineering. NewYork: Springer-Verlag, 2010. 203-297.
- 13 潘明惠.内存计算技术验证项目与应用研究.电力信息与通信技术,2012,(10):44-47.
- 14 奚晓音,何忠江,王丹.基于内存计算技术的电信行业数据仓库系统构建探讨.电信科学,2012,(2) :18-21.
- 15 耿益锋,陈冠诚.Impala:新一代开源大数据分析引擎.http://www.csdn.net/article/2013-12-04/2817707-ImpalaBig-Data-Engine.2013.12.