

文章编号:1001-9081(2006)04-0992-03

一种新的基于神经网络的 IRT 项目参数估计模型

汪存友,余嘉元

(南京师范大学教育科学学院,江苏南京 210097)

(c. y. water@163. com)

摘要:探讨了一种新的基于广义回归神经网络(GRNN)的 IRT(项目反应理论)项目参数估计建模方法,着重介绍了如何建立网络的输出模式及利用 Monte Carlo 方法建立网络的输入模式,提出了多种对模型进行改进的方法。模拟实验表明,利用 GRNN 可以以任意精度拟合 CTT(经典测验理论)参数统计值和 IRT 参数值间隐含的非线性关系。与其他方法进行的比较表明,在小样本情况下,该方法的参数估计误差更小。

关键词:广义回归神经网络;项目反应理论;参数估计;Monte Carlo 方法

中图分类号: TP183;TP391.7 **文献标识码:** A

New model of IRT item parameter estimation based on neural networks

WANG Cun-you, YU Jia-yuan

(School of Education Science, Nanjing Normal University, Nanjing Jiangsu 210097, China)

Abstract: A new modeling method based on general regression neural networks(GRNN) of item parameter estimation within IRT(Item Response Theory) was discussed. The methods about how to construct the output pattern of neural networks, and especially the input pattern by using Monte Carlo method were described. Methods about how to improve the learning efficiency and generalization ability have been proposed. Simulation experiments denote that it is feasible to fit the nonlinear relationship of item parameters between CTT(Classical Test Theory) and IRT given any precision. Comparisons of this method to other methods have been done at last, which suggested it somewhat advantageous.

Key words: general regression neural networks; Item Response Theory (IRT); parameter estimation; Monte Carlo method

0 引言

进行计算机自适应考试(Computerized Adaptive Test, CAT),必须首先建立大型题库。在建立题库前,必须先对项目参数(难度、区分度等)进行估计^[1]。目前国际上流行的项目参数估计软件如 BILOG、MicroCAT 等都是基于项目反应理论(IRT)的统计方法,该方法的缺点是需要大样本量,且计算公式复杂。另一种参数估计理论是经典测验理论(CTT),由于其估计值依赖于特定的测验项目和被试(指参加测验的考生),因而不适合用于题库建设。

文献[2]提出了利用人工神经网络进行项目参数估计的新思路,该方法不仅不需要进行复杂的公式计算,还可以在小样本情况下进行参数估计。文献[2,3]分别利用 RBF 网络和 BP 网络进行了尝试,实验结果表明,在小样本下,神经网络估计方法明显优于 IRT 统计方法。但目前运用神经网络建立参数估计模型的研究还不够成熟,建模的方式并不唯一,如文献[2]以项目反应矩阵为输入、IRT 参数为输出建立神经网络模型,而文献[3]以降维后的矩阵为输入、IRT 参数为输出建模。

1 建模的新思路

因为无论是 CTT,还是 IRT,其项目难度的含义都是题目

的难易程度,区分度的含义都是题目把不同水平的考生区分开来的能力,因此它们之间必然存在某种内隐的关系。洛德经过大量的研究发现,两种方法估计得出的参数之间的确存在一定的非线性关系^[4]。

因此,本文提出一种新的建模方法:通过利用 CTT 方法得到项目参数的 CTT 值,并将之作为网络输入,将 IRT 参数作为网络输出,建立广义回归神经网络(GRNN),通过训练,使网络能够拟合二者间的非线性关系。在实际应用中,就可以将较小样本下 CTT 的项目统计量输入到已经训练好的网络从而获得 IRT 项目参数。

2 模型的理论基础

2.1 IRT 模型

为得到题库的项目参数,通常先编写一份试卷,然后对考生进行预测,得到得分矩阵,通过建立一定的 IRT 模型,并利用统计方法如 JMLE 等^[1]进行迭代计算,最后得到项目参数的估计值。项目反应理论中常用的双参数 logistic 模型为:

$$P_j(\theta_i) = (1 + \exp[-D a_j(\theta_i - b_j)])^{-1} \quad (1)$$

其中, D 取 1.702, θ_i 为考生 i 的能力值, a_j, b_j 分别为项目 j 的区分度值、难度值, $P_j(\theta_i)$ 为考生 i 答对项目 j 的概率。设已知 M 个考生的能力值和 N 个项目的参数值(难度和区分度),

收稿日期:2005-10-18;修订日期:2005-12-12 **基金项目:**全国教育科学“十五”规划项目教育部重点课题资助项目(DBB010511);国家汉办“十五”科研规划资助项目(HBK01-05/088);江苏省教育科学“十五”规划资助项目(D/2001/01/095);江苏省省教育厅自然科学研究资助项目(02KjB180005)

作者简介:汪存友(1982-),男,湖北随州人,硕士研究生,主要研究方向:人工智能、IRT 参数估计;余嘉元(1949-),男,江苏无锡人,教授,博士生导师,主要研究方向:认知心理学、心理测量。

利用公式(1)可以得到一个概率反应矩阵 $(p_{ij})_{M \times N}$ 。当测验为二值记分时(如题目为客观题),我们只能得到一个0(答错)1(答对)组成的得分矩阵 $(x_{ij})_{M \times N}$ 。但由于施测过程考生总会受主客观等随机因素干扰,使得用同一份试卷对同一组考生施测 K 次,每次的得分矩阵 $(x_{ij})_{M \times N}^k (k = 1, 2, \dots, K)$ 并不相同。如果重复施测足够多的次数,然后对所有得分矩阵求和取平均,得到一个新的矩阵,该矩阵各元素的值应该收敛于 $(p_{ij})_{M \times N}$,即:

$$\lim_{K \rightarrow \infty} \left(\sum_{k=1}^K x_{ij} \right) / K = p_{ij} \quad (2)$$

2.2 广义回归神经网络的结构

GRNN(General Regression Neural Networks)是一种正则化的径向基网络,常用来实现函数拟合。它具有一个径向基隐层和一个线性输出层^[5]。基于 Matlab 神经网络工具箱的 GRNN 结构如图 1 所示。

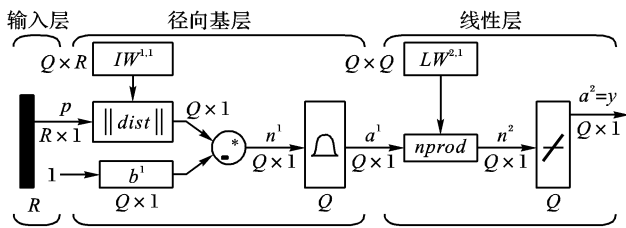


图 1 基于 Matlab 工具箱的 GRNN 结构

其中,网络的隐层和线性层神经元数均等于训练样本数,隐层权值 $IW^{1,1}$ 为训练样本输入的转置,且该层网络输入函数和传递函数分别为 Euclidean 距离函数 $dist()$ 和 Gaussian 函数 $radbas()$ 。线性层权值 $LW^{2,1}$ 为训练样本的输出期望值,该层的网络输入函数和传递函数分别规则化点积函数 $nprod()$ 和线性函数 $purelin()$ 。GRNN 的特点是人为调节的参数少,只有一个宽度值(也称平滑因子);网络的学习全部依赖数据样本;学习速度快;最后收敛于样本量积聚最多的优化回归面。

Matlab 神经网络工具箱提供的广义回归网络设计函数为 $newgrnn(P, T, SP)$ ^[6],其中 P 为输入样本, T 为输出期望, SP 为各径向基神经元的宽度。因此,设计 IRT 项目参数估计的 GRNN 网络,应包括:①设计网络输入模式;②设计网络输出模式;③设计隐层神经元的宽度。

3 基于 GRNN 的 IRT 项目参数估计模型

由于广义回归神经网络是监督性网络,在训练时必须能同时获得网络的输入和输出样本,因此,首先假设项目的 IRT 参数已知(实验中,这些数据都是由计算机模拟产生的),得到网络的输出,再由公式(2)推导出网络的输入。

本研究中,对难度参数和区分度参数的估计分别来建立网络模型,由于两个建模是同步进行的,且方法相仿,为叙述方便,下面仅以难度网络的建模为例。未加特别说明时,各操作在进行区分度估计时同样使用。

3.1 网络输出模式

用计算机模拟产生满足特定概率密度分布的能力参数 $(\theta_i)_M$ 、项目难度参数 $(b_j)_N$ 和区分度参数 $(a_j)_N$ 。其中 M 为考生数, N 为项目数, $i = 1, 2, \dots, M, j = 1, 2, \dots, N$ 。将 $(b_j)_N$ 作为网络的输出,即网络输出设计为维数为 1、样本数为 N 的向量。

3.2 网络输入模式

将 3.1 节中的各参数代入公式(1)得到概率反应矩阵 $(p_{ij})_{M \times N}$ 。用 Monte Carlo 方法模拟施测过程,即设置一个 $[0,$

$1]$ 均匀分布的随机阈值矩阵 $(r_{ij})_{M \times N}$,如果 $p_{ij} \geq r_{ij}$,则 $x_{ij} = 1$, 否则 $x_{ij} = 0$ ^[7],从而得到本次施测的得分矩阵 $(x_{ij})_{M \times N}^k$ 。重复独立施测 K 次,直到公式(3)的值满足指定精度 ε 。

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(p_{ij} - \frac{1}{K} \sum_{k=1}^K x_{ij} \right)^2} \quad (3)$$

其中, $RMSE$ 表示均方根误差。

对于得分矩阵 $(x_{ij})_{M \times N}^k$,采用经典测验理论方法,通过计算各项目的答对率得到 b_j^k (对于区分度,则运用点二列相关法来求 a_j^k),具体方法请参考文献[8]。依次计算 K 个得分矩阵,对于每个项目 j ,可得到一个长度为 K 的向量;如果项目数为 N ,则得到的是一个 $K \times N$ 矩阵 $(b_{kn})_{K \times N}$,将之作为网络的输入模式。

3.3 隐层神经元的宽度

Matlab 提供的 $newgrnn()$ 函数默认的是将所有隐层神经元的宽度设为 1,但显然各个网络训练样本的情况是不同的,使得每个网络都有自己的一个最佳宽度值。常用的寻找最佳宽度的方法是试凑法(try-error-try)。

4 模型拟合能力的验证

用计算机模拟产生一组 $(\theta_i)_M, (a_j)_N, (b_j)_N$,其分布分别为 $[-3, 3]$ 的标准正态分布, $[0, 2]$ 的均匀分布和 $[-3, 3]$ 的标准正态分布。其中 $M = 100, N = 80$ 。按照第 3 节中的方法,获得网络输入输出数据,将网络隐层神经元的宽度设为默认的 1,对网络进行训练,然后利用该训练样本进行仿真,并记录仿真结果的均方根误差。对同一批输出样本,因 Monte Carlo 模拟次数的不同可以得到不同的输入模式,从而训练得到不同的网络。图 2 考察了不同 K 值下训练得到的 GRNN 网络对训练样本自身拟合的能力。

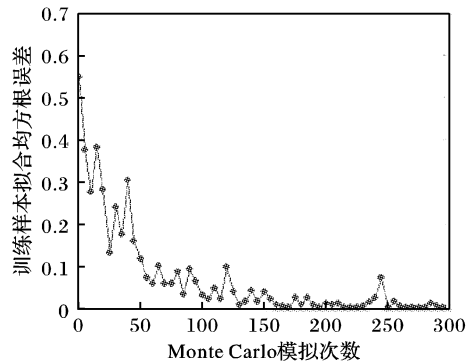


图 2 训练样本拟合误差随模拟次数变化曲线

从图 2 可以看出,当模拟次数达到 150 以上时,利用 GRNN 已经可以很精确地拟合 CTT 参数统计值和 IRT 参数值间的隐含关系,当模拟次数达到 300 时,拟合误差近似为 0。这说明,利用 GRNN 进行 IRT 项目参数估计是可行的。

5 模型的改进及实验结果

一个好的网络必须具有良好的学习性能和泛化能力。基于 GRNN 的 IRT 项目参数估计模型从以下几个方面作了改进:

(1) Monte Carlo 模拟表明,随着实验次数 K 的增加,实验误差将以 \sqrt{K} 倍递减^[9]。但由于 K 值决定了网络输入数据的维数, K 越大,则输入样本维数越高,网络的结构也就越复杂。因而又会影响到网络的泛化能力。所以,当 Monte Carlo 模拟达到预设的精度时,应不再增加 K 值。

(2)对数据进行预处理有利于神经网络的学习^[10]。由于该模型的输入数据范围已经为 $[0,1]$,因此,可以对它进行标准化处理。输出数据范围在 $[-3,3]$ (区分度在 $[0,2]$),可以进行归一化处理。

(3)为寻找网络的最佳宽度,可以用同一批验证数据来检验网络的拟合误差。假设其网络验证误差与隐层神经元宽度间的关系近似一凸函数,可采用黄金分割法来寻找该网络的最佳宽度^[11]。

(4)Perrone 和 Cooper 证明^[12],将神经网络集成运用于回归估计时,如果采用简单平均,且各网络的误差期望为互相独立的随机变量时,则集成的泛化误差为各网络泛化误差平均值的 $1/S$,其中 S 为集成中子网的数目。由于本模型类似于回归估计,所以可以训练多个 GRNN 网络。测试时,网络的最后输出取各网络输出的累加平均。

本研究中,训练样本的 M,N 分别为250,500;验证样本的 M,N 分别为80,100,测试样本的 M 固定为50, N 分别取30、50、80、100、170。Monte Carlo 模拟次数 K 和神经网络集成个数 S 均取30。表1列举了每次测试的均方根误差 RMSE。

表1 相同被试数不同项目数的 RMSE 值

测试项目数	难度 RMSE	区分度 RMSE
30	0.2560	0.1576
50	0.2200	0.1564
80	0.2058	0.1785
100	0.2061	0.1770
170	0.1989	0.1868

6 结语

介绍了一种新的基于 GRNN 的 IRT 项目参数估计建模方法,并结合实际问题对模型进行了多方面的改进,力图利用 GRNN 来学习 CTT 参数统计值和 IRT 参数值之间的隐含关系。实验结果表明:(1)运用 GRNN 进行 IRT 项目参数估计是可行的;(2)在小样本情况下,与现有的 IRT 参数估计软件

相比,该方法的参数估计误差更小^[13];(3)与文献[2]的结果相比仍然存在差距,导致差距的原因除了研究的对象和问题不同外,另一方面,也可能是因为计算机模拟生成的样本质量还不能完全代表实际样本空间,这是值得进一步研究的问题。

参考文献:

- [1] 漆书青,戴海崎,丁树良.现代教育与心理测量学原理[M].南昌:江西教育出版社,1998.313-347,124-177.
- [2] 余嘉元.基于联结主义的连续记分 IRT 模型的项目参数和考生能力估计[J].心理学报,2002,34(5):522-528.
- [3] 谭云兰,丁树良,辛锐铭.基于 IRT 模型的 BP 神经网络估计[J].计算机工程与应用,2004,40(17):56-57.
- [4] 余嘉元.项目反应理论及其应用[M].南京:江苏教育出版社,1992.62-157.
- [5] 董长虹. Matlab 神经网络与应用[M].北京:国防工业出版社,2005.124-126.
- [6] 闻新,周露,李翔,等. Matlab 神经网络仿真与应用[M].北京:科学出版社,2003.285-289.
- [7] LIM RG, DRASGOW F. Evaluation of Two Methods for Estimating Item Response Theory Parameters When Assessing Differential Item Functioning[J]. Journal of Applied Psychology, 1990, 75(2): 164-174.
- [8] 朱德全,宋乃庆.现代教育统计与测评技术[M].重庆:西南师范大学出版社,1998.227-255.
- [9] 高祖新,尹勤.实用统计计算[M].南京:南京大学出版社,1996.130-140.
- [10] 谢庆生,尹健,罗延科.机械工程中的神经网络方法[M].北京:机械工业出版社,2003.88-92.
- [11] 卢险峰.最优化方法应用基础[M].上海:同济大学出版社,2003.67-72.
- [12] PERRONE M, COOPER L. When networks disagree: Ensemble method for neural networks, Artificial Neural Networks for Speech and Vision [M]. New York: Chapman & Hall, 1993. 126-142.
- [13] 谭云兰,丁树良,辛锐铭.基于 IRT 模型的 BP 神经网络降维法参数估计及其应用[J].江西师范大学学报(自然科学版),2004,28(6):485-488.

(上接第 991 页)

3 分布式实时仿真系统高精度时间同步方案

本文设计的分布式实时仿真系统高精度时间同步方案为分层混合时间同步方案。在分布式实时仿真系统的每个局域网内设置一台带有 GPS 时统设备的 PC 机作为时间服务器,局域网时间服务器之间通过 GPS 时统取得同步;局域网内其他计算机通过 LTS 算法与时间服务器取得同步,客户机采用本文设计的新时钟为仿真进程提供时间。客户机性能计数器频率在仿真开始前,利用本文设计的网络测频算法进行测量。图4是仅在仿真开始前进行一次时间对齐,客户机时间与 GPS 标准时间偏差的连续测试图(横轴为时间轴,单位为 s;纵轴为时间偏差轴,单位为 ms)。

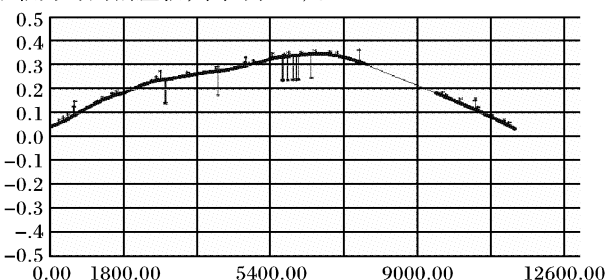


图4 客户机时间偏差曲线

从测试结果看,本方案只在仿真开始前进行一次时间对齐即可保证在3个小时内客户机时间与标准时间的偏差不超过0.4ms。我们在C3I仿真试验系统的时间同步软件开发中将LTS算法的时间对齐间隔设置为30min,连续24h的测试表明时间同步误差最大值不超过0.4ms。

4 结语

本文设计的新时钟无论从时间分辨率方面还是从时间飘移率方面来讲,其优越性都是PC机系统时钟所无法比拟的。本文给出的LTS算法在局域网环境下具有比NTP算法更好的时间对齐效果,并且降低了对网络资源的占用。总之,本文设计的分层混合时间同步方案时间同步精度高,资源占用少,成本低,在保证时间同步精度的前提下,可以将时间对齐对实时仿真运行的冲击降到很低的水平,是分布式实时仿真系统理想的时间同步方案。

参考文献:

- [1] 薛芳侠. C3I 仿真试验系统时间同步技术研究[D]. 国防科技大学, 2004.
- [2] 孙娜,熊伟,丁宇征. 时钟同步的研究与应用[J]. 计算机工程与应用, 2003, 39(27): 177-179.
- [3] 黄四牛,熊新平. DIS 中时钟同步问题研究[J]. 系统仿真学报, 2000, 12(5): 9-10, 17.