

基于神经网络的项目参数估计方法

余嘉元 汪存友

(南京师范大学心理系 南京 210097)

摘要 针对题库建设中项目参数估计的实际问题,提出了一种全新的基于神经网络的参数估计方法;并以二值记分的 3PLM 为项目反应理论模型,以广义回归神经网络为网络模型,根据 Monte Carlo 实验法进行了模拟实验研究,最后将神经网络方法与传统的数理统计估计方法进行了比较。结果表明,在小样本测验情况下,神经网络方法具有一定的优势,尤其是当去掉对项目参数的先验概率分布的限制时,神经网络方法的优势更加明显,说明本文提出的方法具有一定的价值。

关键词 广义回归神经网络,题库,项目反应理论,参数估计,小样本

Method of Item Parameter Estimation Based on Neural Networks

YU Jia-Yuan WANG Cun-You

(Department of Psychology, Nanjing Normal University, Nanjing 210097)

Abstract To solve the problem on item parameter estimation when constructing item bank, a new method based on neural networks (NN) is proposed here. And by choosing three-parameter logistic model (3PLM), general regression neural networks (GRNN) and Monte Carlo as the item response theory (IRT) model, NN model and experimental method respectively, some simulation experiments are conducted. To inspect whether our method works well, a comparison is carried out between the NN method and traditional method, which indicates that the former is more advantageous, especially when the traditional method cancels its item prior constraints.

Keywords General regression neural networks, Item bank, Item response theory, Parameter estimation, Small sample

1 引言

题库中试题(也称为“项目”)参数的预先估计是实施计算机自适应测验(Computerized Adaptive Test, CAT)的重要前提,它的心理测量学基础是项目反应理论(Item Response Theory, IRT)^[1]。传统的项目参数估计方法通常基于数理统计原理,如极大似然法、贝叶斯估计法等等,这类方法的不足在于需要大量考生参加编有新试题的试验性考试(称之为“预测”),不仅增大了题库建设的成本,同时也增加了试题提前曝光的风险。因此,探讨较少考生样本下(以下简称为“小样本”)的 IRT 项目参数估计是非常有价值的。

文[2]提出了一种全新的运用神经网络进行参数估计的方法,它通过将所有考生在项目上的反应模式作为网络输入,将项目的 IRT 参数值作为网络输出,进行网络建模,并通过“锚题”测验设计,提供训练样本以及训练网络;训练好的网络就可以对小样本和连续记分的 IRT 参数进行估计,其模拟实验表明该方法可以得到了较好的估计精度;文[3]对上文进行了改进,消除了网络训练样本对“锚题”的依赖,使得此神经网络估计方法更容易应用于实际测验情景。

但是,文[2,3]的不足在于其所采用的 IRT 模型是连续记分的,在题库建设中并不常用,而基于二值记分的三参数 Logistic 模型(3PLM)则更为实用;所谓二值记分,即指考生在项目上的作答要么正确,要么错误,如:单项选择题、判断题等。那么,适合于连续记分的神经网络项目参数估计方法是否也适用于二值记分方式呢?本文利用广义回归神经网络

(General Regression Neural Networks, GRNN)作为网络工具,并采用 Monte Carlo 实验法^[4],对此问题进行了模拟实验研究。

2 基础理论与模型

2.1 二值记分的 3PLM 模型

项目反应理论把考生在项目上的反应与考生的能力及项目参数之间的关系用一个非线性函数来表示,3PLM 的函数式如公式(1)所示:

$$p_i(\theta) = c_i + (1 - c_i) \frac{e^{D a_i (\theta - b_i)}}{1 + e^{D a_i (\theta - b_i)}} \quad (1)$$

其中, θ 表示所测试的考生的能力, a_i 、 b_i 和 c_i 分别表示项目 i 的基本参数:区分度、难度和猜测系数, D 表示常数值, $p_i(\theta)$ 表示能力为 θ 的考生答对项目 i 的概率。如果以 θ 为横坐标值、 $p_i(\theta)$ 为纵坐标值,并限定考生的能力取值范围在 $[-3, 3]$, 则对于项目 i , 可以绘出一条 S 型曲线。项目 i 的猜测系数 c 为曲线与纵轴交叉处的概率值;区分度 a 的取值与曲线拐点处斜率成正比;难度 b 取纵轴概率为 0.5 时曲线所对应的横轴值。

假设有 M 个考生参加编有 N 道试题的测验,当采用二值记分方式时,若某考生答对项目 i , 则记为 1 分, 否则记为 0 分;那么,对于项目 i , 可以得到长度为 M 的由 0 或 1 组成的反应模式;项目参数估计的目的就在于为这个反应模式找到一组合理的参数 $\hat{a}_i$ 、 $\hat{b}_i$ 、 $\hat{c}_i$ 。

2.2 广义回归神经网络

GRNN 是由 Specht 于 1991 年提出的^[5]。它是一种前馈式监督型神经网络。该网络的结构设计比较简单,它具有一个径向基隐层和一个线性输出层,只有一个网络参数需要调节,具有学习速度快、拟合能力强等优点,常用于解决回归问题,尤其非常成功地应用在函数拟合中。

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* 为径向基隐层神经元的宽度(也称为“平滑系数”)。基于 Matlab 神经网络工具箱的 GRNN 结构如图 1 所示。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 距离函数和 Gaussian 函数。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和线性函数。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。

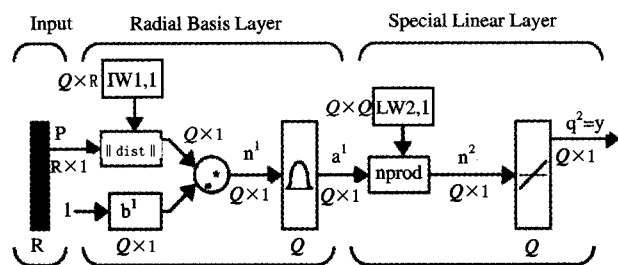


图 1 基于 Matlab 工具箱的 GRNN 结构图

2.3 神经网络建模方法及其理论依据

极大似然法是 IRT 参数估计的主要方法,随后提出的许多参数估计方法也都是由此发展而来的^[7]。所谓极大似然估计,就是为考生在项目上的反应模式建立似然函数,通过迭代计算,找到使反应模式与似然函数最大化匹配的参数,这些参数即为所要估计的 IRT 值。对于项目参数估计,假设通过预测得到一个 $M \times N$ 的反应矩阵, *M* 表示考生数, *N* 表示项目数,每个列矢量表示所有考生在该项目上的反应模式,称为“项目反应模式”,它反映了项目的基本特征;每个反应模式可建立一个似然函数,IRT 项目参数估计的目标就是要找到使此似然函数与该反应模式最大化匹配的参数值。用式(2)表示,即为:

$$L(\bar{y}; \hat{a}, \hat{b}, \hat{c}) = \max_{a, b, c \in \Omega} L(\bar{y}; a, b, c) \quad (2)$$

其中, $\bar{y}$ 表示所有考生在项目上的反应模式, *L* 表示反应模式的似然函数, Ω 为项目所有可能取值的集合,当参数组合 $\hat{a}_i$ 、 $\hat{b}_i$ 、 $\hat{c}_i$ 使得 *L* 取值最大时,其即为该项目的 IRT 参数值。

当项目上的反应模式不同时,其估计出的参数值也可能不同,例如,反应模式分别为 $[1 \ 0 \ 1]^T$ 和 $[1 \ 1 \ 0]^T$ 的两道试题的项目参数很可能并不相同;可见,在反应模式和 IRT 项目参数之间是存在某种关系的。倘若能够提供反映了项目反应模式与项目参数间关系的学习样本,用它们来训练神经网络,使之学习到二者间的关系,那么,就可以将参数未知的项目反应模式输入到网络,从而输出其 IRT 参数估计值了。

3 实验研究

3.1 实验数据的产生办法

考生在项目上的反应模式可通过三个步骤模拟得到:首先,将模拟得到的能力参数 θ_j 和项目参数(a_i 、 b_i 、 c_i)代入公式

(1)后得到反应的几率值 p_{ij} ,然后,产生一个 $[0, 1]$ 范围内矩形分布的随机数 *r*;最后,若 $p_{ij} \geq r$,则判定考生 *j* 答对项目 *i*,否则判定为答错。将考生在项目上的反应模式作为神经网络的输入,将该项目的 IRT 参数值作为神经网络的输出,即可完成神经网络的建模。能力参数和项目参数的取值范围、概率分布及其产生函数如表 1 所示。

表 1 IRT 参数总体分布及取值范围

IRT 参数	总体分布	MATLAB 函数	取值范围
能力 θ	标准正态	<i>randn</i> ()	$[-4.4174, 4.4174]$
难度 <i>b</i>	标准正态	<i>randn</i> ()	$[-3.1073, 3.1073]$
区分度 <i>a</i>	对数正态	<i>log(randn())</i>	$[0.3781, 1.7728]$
猜测系数 <i>c</i>	beta 分布	<i>beta</i> (3,30)	$[0.0061, 0.3042]$

3.2 实验设计及步骤

实验中,要求预先提供部分 IRT 参数已知的项目,并将之同待估计的测验项目编制在一份试卷上,参数已知的那部分项目可以作为网络的训练样本,而待估计的项目则可作为测试样本。要完成对待估计的项目参数值的估计,需按如下步骤进行,在此过程中将分别建立一个能力参数估计神经网络和项目参数估计神经网络:

Step1 模拟产生 M_1 个考生能力参数(称为“参加预测的考生”);模拟产生 N_1 个项目 IRT 参数(称为“已知参数项目”);然后,根据 3.1 节中的方法,产生模拟的考生反应矩阵(称为“反应矩阵 I”);

Step2 模拟产生 N_2 个项目 IRT 参数(称为“待估项目”);然后,根据 3.1 节中的方法,将之与“参加预测的考生”反应,产生模拟的考生反应矩阵(称为“反应矩阵 II”);

Step3 模拟产生 M_2 个考生能力参数(称为“模拟考生”);然后,根据 3.1 节中的方法,将之与“已知参数项目”反应,产生模拟的考生反应矩阵(称为“反应矩阵 III”);

Step4 根据“反应矩阵 III”和“模拟考生”按行模式建立并训练能力参数估计神经网络;GRNN 的平滑系数采用凑试法确定;

Step5 将“反应矩阵 I”按行模式输入到已经训练好的能力参数估计神经网络中,得到“参加预测的考生”的能力参数的神经网络估计值(称为“已估计考生”);

Step6 模拟产生 N_3 个项目 IRT 参数(称为“模拟项目”);然后,根据 3.1 节中的方法,将之与“已估计考生”反应,产生模拟的考生反应矩阵(称为“反应矩阵 IV”);

Step7 根据“反应矩阵 IV”和“模拟项目”按列模式建立并训练项目参数估计神经网络;GRNN 的平滑系数采用凑试法确定;

Step8 将“反应 II”按列模式输入步骤 7 中已经训练好的项目参数估计神经网络中,得到“待估项目”的 IRT 参数的神经网络估计值。

为了减小模拟实验中的随机误差,对项目参数重复抽样 R_1 次,并在固定一批项目参数的情况下,对考生能力参数进行重复抽样 R_2 次。

3.3 误差统计指标

本文采用 MC 方法产生神经网络的训练样本和测试样本,由于网络的输出也是已知的,因此,可以直接将神经网络的项目参数估计值与模拟的真实值进行比较,从而确定神经网络的估计误差。本文设计了三种误差统计指标,分别是:平均绝对偏差(Mean Absolute Bias, MAB),均方根误差(Root

Mean Squared Error, RMSE) 和相关系数 (Correlation Coefficient, COEF)。相关系数的计算公式为皮尔逊积差相关系数, MAB 和 RMSE 的计算公式分别为:

$$MAB = \frac{1}{N_2} \sum_{i=1}^{N_2} \frac{1}{R_2} \left| \sum_{r=1}^{R_2} (x_i - \hat{x}_{ir}) \right| \quad (2)$$

$$RMSE = \frac{1}{N_2} \sum_{i=1}^{N_2} \sqrt{\frac{1}{R_2} \sum_{r=1}^{R_2} (x_i - \hat{x}_{ir})^2} \quad (3)$$

其中, N_2 表示“待估项目”数, R_2 表示考生重复抽样数, r 表示对考生进行第 r 次重复抽样, x_i 和 $\hat{x}_{ir}$ 分别表示第 i 个项目的真实值和第 r 次考生抽样下的估计值。由于 COEF 仅仅反映了估计值与真实值间的一致性程度, 其对误差的反映程度最弱, 因此只作为参考指标。

表 2 神经网络与 BILOG-MG 方法的估计误差比较

	MAB		RMSE		COEF	
	GRNN	BILOG-MG	GRNN	BILOG-MG	GRNN	BILOG-MG
区分度	0.1926/0.0320	0.3570/0.0305	0.2397/0.0297	0.4846/0.0308	0.3978/0.0725	0.5823/0.0612
难度	0.1824/0.0457	0.1275/0.0273	0.3230/0.0423	0.3519/0.0759	0.9758/0.0068	0.9680/0.0182
猜测系数	0.0299/0.0080	0.0502/0.0104	0.0418/0.0106	0.0691/0.0126	0.6773/0.0543	0.1202/0.0420

注: “/”前数字表示 30 次重复实验误差的平均值; “/”后数字表示 30 次重复实验误差的标准差。

特别需要说明的是, BILOG-MG 的主要算法是基于贝叶斯方法, 它需要对估计对象的先验概率分布做出限定^[9]。本文实验中, 还对 BILOG-MG 程序做了相关修改, 去掉了其对项目先验概率的限定, 在“参加预测的考生”数为 500、“待估项目”数为 20 的情况下, 得到的估计结果如表 3。

表 3 去掉先验概率密度限时 BILOG-MG 方法的估计误差

	MAB		RMSE		COEF	
	M	SD	M	SD	M	SD
区分度	0.5905	0.0181	0.6579	0.0221	0.0758	0.1019
难度	0.3447	0.0323	0.3912	0.0274	0.9918	0.0021
猜测系数	0.1283	0.0236	0.1640	0.0241	0.0384	0.1450

注: M 表示 30 次重复实验误差的平均值; SD 表示 30 次重复实验误差的标准差。

4.2 分析及讨论

分析表 2 中的 MAB 指标可以发现, 在区分度和猜测系数上, 神经网络方法比数理统计方法的估计误差要小, 但在难度上, 神经网络方法的误差较大; 进一步分析 COEF 指标可以发现, 神经网络方法在难度估计上的相关系数并不差。

分析表 2 中的 RMSE 指标可以发现, 在区分度、难度和猜测系数上, 神经网络方法比数理统计方法的估计误差都要小, 尤其是在区分度和猜测系数上, 神经网络方法的估计精度较明显。

虽然从表 2 的总体结果来看, 神经网络方法的优势并不明显, 但通过分析表 3 可以发现, 当去掉对项目参数的先验概率分布的限制后, BILOG-MG 的估计性能下降得十分厉害, 即使在“参加预测的考生”数达到 500 时, 仍然具有较大的估计误差。可见, 神经网络方法还是具有一定的优势的。

同时, 本文还将神经网络方法同数理统计方法的参数估计性能进行了比较研究, 其中, 数理统计参数估计方法是通过国际上流行的 IRT 参数估计软件 BILOG-MG 来完成的^[8]。

4 实验结果与分析

4.1 实验结果

本文实验中, “参加预测的考生”数 M_1 、“已知项目数” N_1 、“待估项目数” N_2 、“模拟考生”数 M_2 、“模拟项目”数 N_3 、项目重复抽样次数 R_1 、考生重复抽样次数 R_2 分别取 100、15、20、250、200、30、50。实验中 GRNN 网络的平滑系数通过凑试法确认, 默认为 1。实验结果如表 2 所示。

结论 本文针对题库建设中试题参数预测的实际问题, 采用一种全新的基于神经网络的参数估计方法, 并以二值记分的 3PLM 为 IRT 模型, 以 GRNN 为神经网络模型, 根据 Monte Carlo 实验法进行了模拟实验研究, 最后将神经网络方法与传统数理统计估计方法进行了比较。结果表明, 在小样本测验情况下, 神经网络方法具有一定的优势, 尤其是当去掉对项目参数的先验概率分布的限制时, 神经网络方法的优势更加明显, 说明本文提出的方法具有一定的价值。

参考文献

- 漆书清, 戴海崎, 丁树良. 现代教育与心理测量学原理[M]. 北京: 高等教育出版社, 2002. 239~244
- 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
- 余嘉元, 陈淑燕. 运用径向基网络估计项目反应模型参数的研究[J]. 中国考试(考试研究版), 2005, 4: 24~26
- 高祖新, 尹勤. 实用统计计算[M]. 南京: 南京大学出版社, 1996. 130~140
- Specht D. A General Regression Neural Network [J]. IEEE Transactions on Neural Networks, 1991, 2 (6): 568~576
- 董长虹. Matlab 神经网络与应用[M]. 北京: 国防工业出版社, 2005. 124~126
- Baker F. Item Response theory: Parameter Estimation Techniques [M]. New York: Marcel Dekker, 1992
- Zimowski M, Muraki E, Mislevy R, et al. BILOG-MG: Multiple-group IRT Analysis and Test Maintenance for Binary Items [M]. Chicago: Scientific International, Inc, 1996
- Miskey R. Bayes Modal Estimation in Item Response Models [J]. Psychometrika, 1986, 51: 177~195