

应用神经网络方法仿真生育率结构

龚跃 杨华民

摘要 利用BP神经网络方法对人口按龄生育率问题进行建模和仿真,得到了较为理想的拟合结果。这一结果表明使用神经网络方法模拟生育率结构动态具有建模简单和拟合精度高的特点,在降低模型复杂性上优于传统方法。

作者 龚跃,男,1983年毕业于吉林大学数学系,现任长春光学精密机械学院计算机系副教授。(长春 130022)

杨华民,男,1981年毕业于大连理工大学计算机系,1995年毕业于吉林工业大学计算机系,获硕士学位,现任长春光学精密机械学院计算机系副主任、副教授。(长春 130022)

一、引言

生育率作为人口系统的主要动因是人口统计学家所关注的焦点,而按龄生育率又是近年来这一研究方向最活跃的领域^[1]。到目前为止,在生育率模型方面已经取得了一些重要的成果,Henry于1953年提出一种生育率模型^[2],把随机过程理论和卷积方法引入生育率模型构造之中,为后期生育率模型发展奠定了基础。Coale在1972年用一种双指数函数成功地拟合了首次婚姻的年龄分布,提出了著名的Coale-McNeil模型^[3],这一模型被广泛用于生育率建模之中。Sheps1964年把人类再生过程定义为马尔可夫更新过程。从此马尔可夫随机过程模型在生育率建模中得到广泛应用。1983年Mode把随机过程的卷积理论应用到生育率分析之中,即N个随机变量之和的联合分布可表为这N个随机变量原始分布的N重卷积。Jun Zhu 1994年再次完善了生育过程马尔可夫模型。把首次结婚分布引入到生育率模型之中,分别采用Coale-McNeil模型、正态分布函数和指数函数作为原始分布,用卷积构成整个过程。值得一提的是样条技术、卷积最小二乘及快速傅里叶变换等复杂数学工具被用于模型构造和计算之中。这一工作代表着生育率模型的最新进展。

随着研究的深入,一些颇复杂的数学工具,如正态分布、指数分布和双指数分布、最小二乘和卷积方法已被普遍用于生育率建模之中,快速傅氏变换和样条技术也逐步被采用。这一复杂趋势一方面使模型更加逼近实际出生率变化,另一方面也使模型的应用产生困难,如卷积最小二乘技术的复杂程度就是一般人口统计学者所难以接受的。

神经网络模拟了人脑神经元的基本功能和网络的基本结构特征,具有分布式记忆、并行处理、联想、自组织学习能力。利用神经网络良好的非线性映射能力,可以避免复杂的参数估计过程去解决一系列变量关系不能精确地用函数表达的分类与回归问题。同时具体应用时神经网络又兼有概念直观、公式简洁、标准化编程易实现等优点。它为人口统计研究提供了新的方法。

本文使用神经网络方法对非线性时间序列按龄生育率进行建模和仿真取得了令人满意的结果。

二、BP神经网络模型原理

Cybenko^[4]和Fanahashi^[5]从理论上证明具有一个隐层(假设具有足够的隐层节点)的神经网络能够以任意精度表示任何连续函数。因此统计学中复杂的多元非线性回归问题可以通过神经网络来解决^[6]。

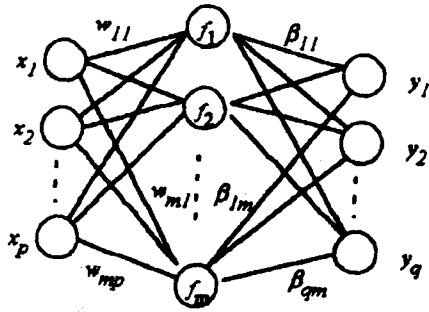


图1 前馈神经网络-BP网络模型

一个多元非参数的回归问题可描述如下:

给定 n 对向量:

$$(Y_L, X_L) = (y_{L1}, y_{L2}, \dots, y_{Lq}; x_{L1}, x_{L2}, x_{Lp}); \quad L = 1, 2, \dots, n \quad (1)$$

有未知数学模型:

$$y_{Li} = g_i(X_L) + \varepsilon_{Li}, \quad L = 1, 2, \dots, n; \quad i = 1, 2, \dots, q \quad (2)$$

其中 $|Y_L|$ 称为多元响应向量, $|X_L|$ 称为独立变量。 $|g_i|$ 是未知的平滑非参数函数, g_i 完成从 P 维 Euclidean 空间到实线的映射:

$$g_i: R^P \rightarrow R; \quad i = 1, 2, \dots, q$$

(2)式中 $\{\varepsilon_{Li}\}$ 是均值为零的随机变量, $E[\varepsilon_{Li}] = 0$, $\{\varepsilon_{Li}\}$ 也常常被假定为独立变量。

回归分析的目标是构造估计: g_1, g_2, \dots, g_q , 以便更好地逼近未知函数: g_1, g_2, \dots, g_q , 然后使用这些估计对新给出的 X 预测新的 y 值。

$$y_i = g_i(X), \quad i = 1, 2, \dots, q$$

针对按龄生育率问题, 本文选用典型的前馈神经网络-BP网络模型(见图1),

输出 y_i 的估计为:

$$\begin{aligned} \hat{y}_i &= E[y_i | x_1, x_2, \dots, x_p] \\ &= \sum_{k=1}^m \beta_{ik} f_k \left(\sum_{j=1}^p w_{kj} \cdot x_j - w_{k0} \right) \\ &= \sum_{k=1}^m \beta_{ik} \sigma \left(\sum_{j=1}^p w_{kj} \cdot x_j - w_{k0} \right), \quad i = 1, 2, \dots, q \end{aligned} \quad (3)$$

其中: f_k 是非线性作用函数, 通常选取单调增加的 S 型映射:

$$\sigma(z) = 1/(1 + e^{-z}). \quad (3) \text{式可改写为:}$$

$$\hat{y}_i = g_i(X) = \sum_{k=1}^m \beta_{ik} \sigma(W_k^T \cdot X - w_{k0}), \quad i = 1, 2, \dots, q \quad (4)$$

其中: W_k 是权重向量, X 是输入向量。(4)式说明每当给出一个 X 便可以预估出新的输出 Y 。

三、生育率 BP 网络建模

对于时间序列预测, 假设有历史观测值 $X^n, X^{n-1}, \dots, X^{n-m+1}$, 对未来 $n+k$ 时刻进行预测, 则(4)式中 $Y = X^{n+k}$, $X = X^n, X^{n-1}, \dots, X^{n-m+1}$ 。根据文献^[7], 图1所示网络可以进行一步预测和多步预测。对荷兰国家1980年—1992年, 14岁—51岁年龄间不同胎次的按龄生育率统计数据(以下简称 xx 年数据, 参见表1), 使用一个三层 38:10:38 的 BP 网络, 分别对不同胎次进行一步预测和多步迭代预测实验, 不同胎次生育率预测过程相同。

(1)一步预测建模

不同胎次生育率预测过程相同,取所有输入都是实际观测值,且 $k=1$ 时所做的预测是一步预测。我们先以 1980 年的数据作为训练样本数据,以 1981 年的数据作为训练样本数据,训练网络,建立对应关系;依次取 1981 年数据对应 1982 年,……取 1989 年数据对应 1990 年,这样在 BP 网络上建立了 1980 年—1989 年到 1981 年—1990 年的对应关系。训练后的网络算法从原始数据中自动抽取了生育率变化的规律,替代了数学建模过程。预测 1991 年按龄出生率,结果预测偏差为 0.02901,见图 2。从计算结果可以看出人工神经网络具有较为理想的拟合功能,其一步预测结果也具有较高的精度。

(2)迭代预测

采用迭代一步预测方法,以 1980 年—1988 年数据作训练样本数据,1981 年—1989 年数据作教师数据,训练网络。先以 1989 年数据为输入,预测出 1990 年按龄出生率,然后将该数据反馈回网络输入端再进行 1991 年按龄出生率的预测,1991 年统计数据作检验数据使用。预测结果偏差为 0.04131,见图 3。

四、结 论

在描述生育过程中,传统的数学模型方法采用许多复杂的数学工具使得这一方法越来越难以广泛接受。复杂表达形式导致了超负荷的计算量与计算精度损失。按龄生育率这样多因子的特征指标致使表示它的模型复杂化,而简化模型的工作极为困难。理论分析和预测过程及结果表明,采用人工神经网络非线性时间序列预测方法对按龄生育率预测,具有建模简单、拟合精度高和运算速度快的特点。在预测精度、降低模型复杂性上优于传统方法。

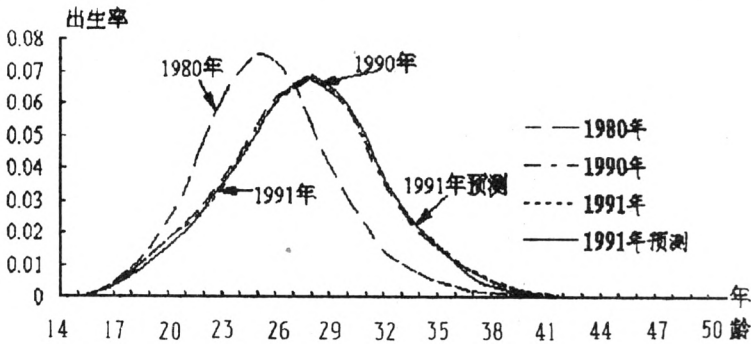


图 2

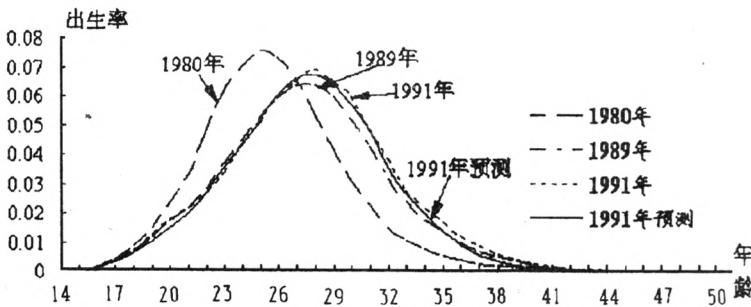


图 3

表1 荷兰1980年14岁—51岁年龄间不同胎次的按龄生育率统计数据

年龄	胎 次				
	1	2	3	4	5
14	0.0000926	0	0	0	0
15	0.0003465	0	0	0	0
16	0.0015375	0.000008	0	0	0
17	0.0044849	0.0002028	0	0	0
18	0.0092281	0.0008270	0.0000246	0.0000082	0
19	0.0148786	0.0024323	0.0001627	0	0
20	0.0237059	0.0058888	0.0005103	0.0000502	0
21	0.0332615	0.0098057	0.0009260	0.0001323	0.0000083
22	0.0486596	0.0172068	0.0020049	0.0002538	0.0000592
23	0.0614602	0.0273915	0.0035862	0.0005294	0.0000939
24	0.0699533	0.0390139	0.0066725	0.0009779	0.000225
25	0.0755778	0.0503039	0.0095278	0.0014631	0.000418
26	0.0730526	0.0609991	0.0138522	0.0023454	0.0004673
27	0.0661039	0.0652340	0.0171762	0.0031246	0.0009143
28	0.0531530	0.0654232	0.0213425	0.0038162	0.0014311
29	0.0412967	0.0605346	0.0233195	0.0044625	0.0015782
30	0.0305995	0.0517748	0.0242506	0.0055103	0.0025973
31	0.0224625	0.0404006	0.0224536	0.0056488	0.002364
32	0.0146705	0.0298551	0.0202147	0.0053043	0.0025536
33	0.0105106	0.0210533	0.0161237	0.0051387	0.0026297
34	0.0074053	0.0149905	0.0123396	0.0043947	0.0025023
35	0.0053892	0.0096314	0.0090037	0.0042097	0.0034305
36	0.0037439	0.0070764	0.0066547	0.0031474	0.0033942
37	0.0022943	0.0043831	0.004567	0.0029004	0.0036147
38	0.0022458	0.0031394	0.0032453	0.0023634	0.003645
39	0.0013101	0.0020569	0.00191	0.0016039	0.0025344
40	0.0009989	0.0015697	0.0013318	0.0010108	0.0032701
41	0.0004623	0.0009733	0.0008516	0.000876	0.0020439
42	0.0004345	0.0005462	0.0005586	0.0005959	0.0016635
43	0.0003008	0.0002093	0.0003139	0.00034	0.0012294
44	0.0001831	0.0000785	0.0001439	0.0001962	0.000641
45	0.0000664	0.0000664	0.0000398	0.0001195	0.0005178
46	0.0000398	0.0000664	0.0000398	0.0001063	0.0003321
47	0.0000404	0.0000404	0.0000135	0.0000674	0.0001482
48	0.000013	0.000026	0.000026	0.000013	0.0000651
49	0	0.0000395	0.0000264	0	0.000145
50	0.0000647	0.0000129	0.0000129	0.0000129	0.0001034
51	0.0000134	0.0000134	0	0	0.0000535

(参考文献见第44页)

育、保险等各种手续。同时, 培训班还与学文化、学科技结合起来, 扫除青壮年文盲, 提高群众的科技文化素质, 使她们掌握科技致富的实用技术, 为发展致富项目打下基础。

四

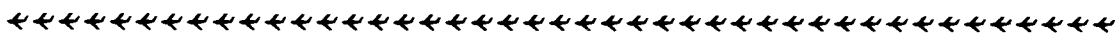
通过对以上调查结果的描述和分析, 我们得到如下几点启示:

1. 杜绝和减少计划外出生, 控制人口过快增长是计划生育的一个重要内容和直接目标, 但是, 按照全面意义上的计划生育要求, 实现了控制人口过快增长只是迈出了第一步, 要真正实现提高人口素质、优化人口结构、达到人口与社会经济协调发展的宏伟目标, 使计划生育工作提高到一个新的水平, 最终必须落实到群众生育观念的转变上, 如果没有广大群众对计划生育认识的提高、观念的转变, 就不可能把党和国家对计划生育的要求变为他们自觉自愿的行动, 即便是靠行政制约实现了控制人口过快增长也是暂时的, 在工作力度上稍有放松, 就还会出现反复。

2. 我们的调查表明, 潮泉镇农民的生育观念确实发生了很大变化, 可以说已经开始跨进现代生育观念的门槛。但是同时我们也认识到, 这种转变还不是或主要不是自发性的, 即不是在经济文化发达条件下完全自发形成的, 而是行政制约、利益引导和

宣传教育等多种因素共同促进引导的结果。然而, 令我们感到欣慰和鼓舞的是, 在经济文化条件相对比较落后的条件下, 潮泉镇农民的生育观念能达到这种程度, 除了说明农民的觉悟提高及党和政府多年来的艰苦工作取得了成效外, 更重要的是说明在落后的经济文化条件下可以形成先进的生育观念并在其引导下使计划生育工作真正出成绩、上水平。

3. 目前, 我国计划生育的不平衡现象不仅存在于不同省市、地市和县市等宏观地域之间, 而且存在于不同乡镇以至不同村落等微观地域之间。为什么在相同的国家政策之下, 有相当一部分地区的经济和文化发展水平都不错, 但其农民的生育观念却相当落后? 为什么在相同的经济和文化发展条件下, 不同地区的计划生育工作却有较大差别? 我们认为, 这和不同的社区环境具有重要关系。因为人是环境的动物, 人的观念和行无不受周围环境的影响。在大小远近各不相同的环境层次中, 对人影响最大和最直接的往往是他直接所处的社区环境, 即通常所说的“小环境”。因此, 在“对人口实行计划生育是我们的一个基本国策”这个相同的大环境下, 对于不同地区和部门来说, 做好计划生育工作、转变人们生育观念的关键在于努力营造一个良好的社区环境。而且, 潮泉镇的经验告诉我们, 只要认识对路、措施得力, 即使在经济文化都不十分发达的条件下, 这种社区环境也是可以营造出来的。



(上接第 53 页)

参考文献

- 1 Jun Zhu. A model of the age patterns of births by parity in natural fertility populations. *Mathematical population studies*. 1994, 2: 153 - 173
- 2 Henry, Louis. *Theoretical basis of measures of natural fertility. On the Measurement of Human Fertility*. New York: Elsevier, 1953
- 3 Coale, A. J., and McNeil, D. R. The distribution by age at first marriage in a female cohort. *Journal of the American Statistical Association* 67(340): 743 - 749, 1972
- 4 Cybenko G. Approximations by Superpositions of sigmoidal Function. *Mathematics of control. Signals and Systems*. 1989(2): 303 - 314
- 5 Fanahashi K. On the Approximate Realization of Continuous Mapping by Neural Networks. *Neural Networks*. 1989, 2: 183 - 192
- 6 J - N. Hwang, M. Maechler. Regression modeling in back - propagation and projection pursuit learning. *IEEE transactions on neural networks*, Vol. 5, No. 3, May 1994: 343 - 353
- 7 Weigend A S, Rumeihart D E, Huberman B A. Back-propagation, weightelimination and time seris prediction. In: *Proceedings of the Connectionist Models*, 1990, 105 - 116