混沌、非线性理论与历史研究

马未然
北京大学信息科学技术学院

摘要： 本文通过研究人口变化来展示非线性理论在历史研究中的作用。文章介绍了较为通用的 Logistic 函数，并具体建模处理人口数据，后又通过虚拟“历史游戏”推演展示更多混沌现象。

关键字： Logistic 函数；Logistic 映射；人口变化；混沌现象

一 引言

混沌系统有几大特征：

1. 系统的演化不确定、类随机。
2. 无法对后续演化做出长期预报。
3. 对初值敏感。
4. 部分系统会出现类似周期变化的现象。

非常巧合的是，在人类历史的发展中，如果将一个较为孤立的政治体看作系统，分析其各项衡量指标随时间的变化，则其也有对应的性质：

1. 许多重大变革或者自然灾害的不定时发生，使得历史发展有极强的不确定性。
2. 无法根据历史去预测未来，尤其是长久以后的未来。
3. 不同的地理环境孕育出截然不同的文明。
4. 在古代，往往会出现“王朝周期律”，历史总是螺旋上升。

这些性质，启发了笔者去探究历史与混沌现象的关系。然而历史过于复杂，难以面面俱到地进行研究分析，故本文只从人口侧面进行研究。

之所以选择人口，是因为其具有在短期内人口的增长速度较为平缓，与其它因素耦合较小的特点。而其它的指标如 GDP 或发展程度等，要么与其它因素耦合较多难以分离，要么变化规律更为随机，研究起来都会遇到很大困难。

由于笔者本身并不是历史专业，故本文会更侧重运用计算机对演化进行数值模拟计算，而非对模型深层原理的探究，这也导致了文章的研究方法、所得结论不免存在很大局限。

二 Logistic 函数与 Logistic 映射

在这篇原始论文[1]中，Verhulst 最早将 Logistic 函数引入社会科学的研究。由于原论文语言并非英文，故研究主要参考此处[2]的解读。

Logistic 函数形如：

$$P(t) = \frac{P(0)e^{rt}}{1+\frac{P(0)(e^{rt}-1)}{K}} = \frac{K}{1+\left(\frac{K}{P(0)}-1\right)e^{-rt}} \tag{1}$$

这个函数满足：

$$\frac{\mathrm{d}P}{\mathrm{d}t} = rP\left(1-\frac{P}{K}\right) \tag{2}$$

事实上，在原始的研究过程中，是先猜测变化规律满足 式 2 的形式，再解出 式 1。

先来定性分析这个函数，主要研究其初值 $P(0)$ 位于 $(0, K)$ 内的性质。不难发现这个函数随 t 单调递增，且不断逼近却永不到达 $P = K$。这样看来，系统尽管非线性，却不存在混沌现象。

其实，对于 式 2 基本现象的分析，并不需要解出 式 1。式 2 是一维自治系统，有不动点 0 和 K。$r > 0$ 时，求二阶导，可知 0 是不稳定平衡点，K 是稳定平衡点，故系统会从 0^+ 向 K^- 演化（这也可通过 $(0, K)$ 中导数非负看出）。

有趣的是，假如我们将数据离散化，那么有：

$$P_{t+1} - P_t = rP_t\left(1-\frac{P_t}{K}\right) \tag{3}$$

也即：

$$P_{t+1} = (r+1)P_t - \frac{P_t^2}{K} \tag{4}$$

变换形式，我们令 $P_t = (r+1)KQ_t$，那么有：

$$Q_{t+1} = (r+1)Q_t - (r+1)Q_t^2 = (r+1)Q_t(1-Q_t) \tag{5}$$

正好是 Logistic 映射的形式！

这样看来，当 r 值较大时，系统演化会出现混沌现象，与之前的分析不同。这是因为，当 r 较大时，离散化选的时间间隔 1 便过于粗糙，无法拟合连续的函数。

实际上，r 代表了人口增长率，一般而言有 $r \sim 10^{-2}$，那么混沌也就无从存在了。

我们注意到，对于较小的封闭生态系统中的昆虫数量，其往往会出现混沌现象，它为什么与人口的性质不同呢？

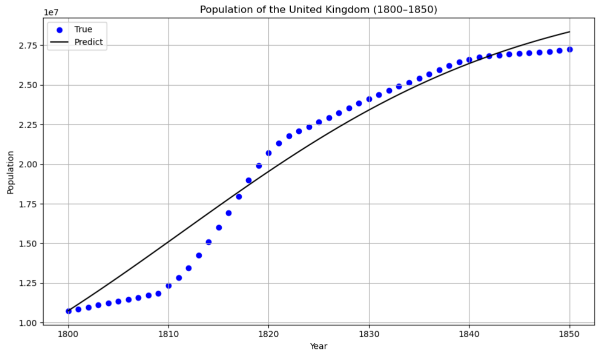
对于昆虫之类的生物，它们在一年中的特定时间进行繁殖，故应采用离散而非连续的方法研究，那么，用 Logistic 映射模型就能较好地拟合其数量发展状况。同时，其增长率 r 往往较大，因此会出现混沌。

三　对于人口数据的建模

现在，我们用 Logistic 映射模型对人口数量进行拟合[1]。

这里，我们选则 $1800 \sim 1850$ 间英国人口变化为研究对象，原因是这段时间中英国社会没有经历重大变革，人口结构相对稳定，数据相对齐全。

我们假设 1800 年时 $t = 0$，那么枚举所有的 r 和 K，发现在 $r = 0.0576$ 及 $K = 31390000$ 时，预测达到的 r2 score[2] 最大，曲线如下：

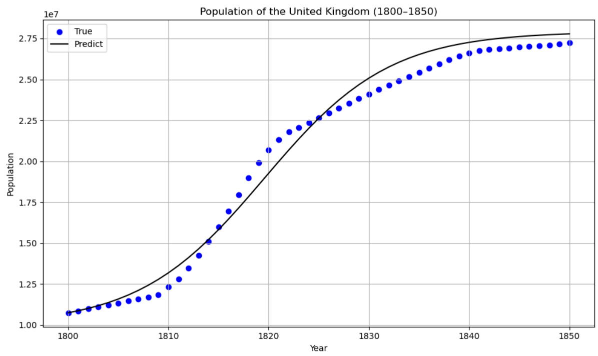


由图可知，其对于 $1800 \sim 1820$ 中的数据拟合不算出色，在 $1820 \sim 1850$ 中则较好地进行了拟合。同时，也看出最优的拟合曲线接近于直线，因为其导数在区间内变化幅度不大。

为了更好拟合曲线，我们将人口整体偏移 10^7，将公式改写成：

$$P(t) = \frac{K}{1 + \left(\frac{K}{P(0) - 10^7} - 1\right)e^{-rt}} + 10^7 \tag{6}$$

这种情况下，最优的 $r = 0.16$，$K = 17250000$，曲线如下：



尽管其更接近真实曲线，但参数 10^7 是根据真实曲线选取的，未免有“过拟合”嫌疑。

[1]数据来自：https://ourworldindata.org/grapher/population

[2]https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html

我们再来讨论 Logistic 函数统计人口时的意义：r 代表了增长率，K 代表了最多能容纳的人口数量。当人口适中时，增长速率最快；而人口过少或过多都会减缓增长速率。

这样的定义整体上是很合理的，但我认为其也存在不合理之处：增长率关于人口数达到 $\frac{K}{2}$ 对称，这个推断未免有些主观。

我们对模型的调整（将人口量整体偏移）实际上就是改变了对称点，改变后，其更符合实际。

更深的不合理之处是，人口并不是孤立的历史变量：

1. 它与经济发展、地缘政治等因素有较强的联系，单独分离出来研究难免失真。
2. 一些偶发的随机现象（如灾难、疫病等），会大大减少人口，这在公式里没有体现。
3. 一些技术革新如农业革命、科技革命等，会大幅改变人口增长率 r 与容纳上限 K。而在模型中，r 和 K 都是不变量，与实际不符。

所以，Logistic 函数在处理稳定社会的短期人口数据时效果很好，可是对于长期情况或处于剧烈变化中的社会便束手无策了。

四　虚拟“历史游戏”推演

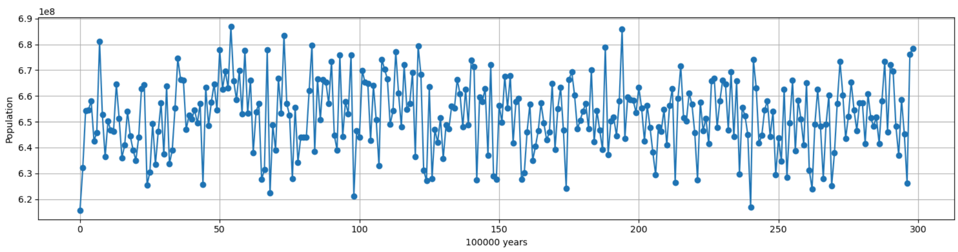
对于真实的历史演化情况，难以用几个参数就完全描述。在这里，笔者使用计算机建模的方法，设计了一个虚拟的“历史游戏”，来进行模拟演化并分析其人口变化。

在“历史游戏”中，笔者设计了 11×11 个国家，记录每个国家的人口，计算总人口。人口按照 Logistic 映射规则，并且其 r 和 K 是可变的，会以小概率随机增减。

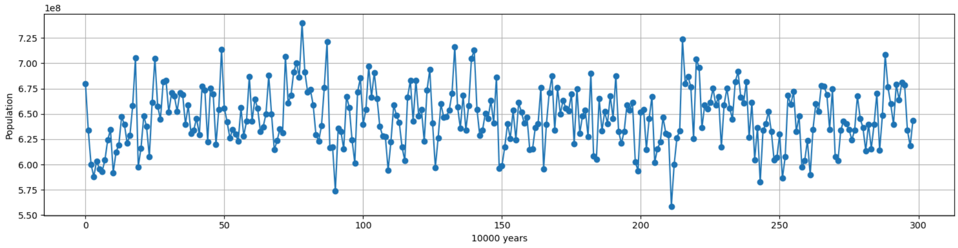
同时，为模拟不同国家间人口的关联，笔者设计了“移民”：每个国家每年会有一定人口向周围更“移居”的国家移民。移居指数由 r、K 和人口数 p 决定。

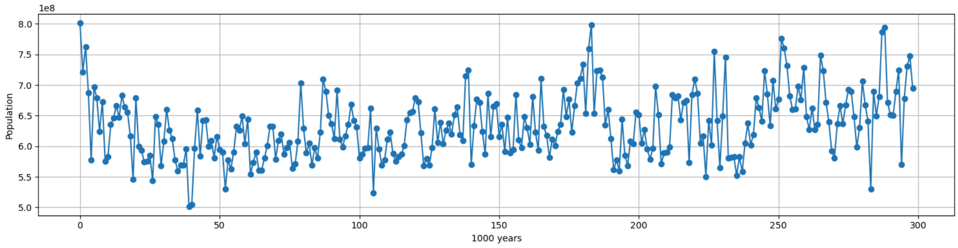
具体细节在此处不再赘述，可参考代码 game.cpp[3]。

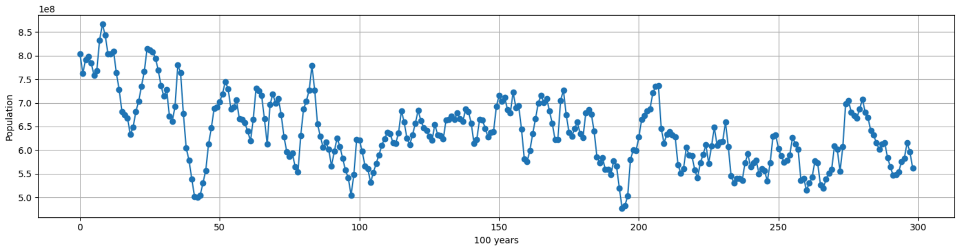
这里展示了总人口随时间的变化（对某个时间周期内的人口数求平均值），时间尺度由 10^5 年至一年不等。

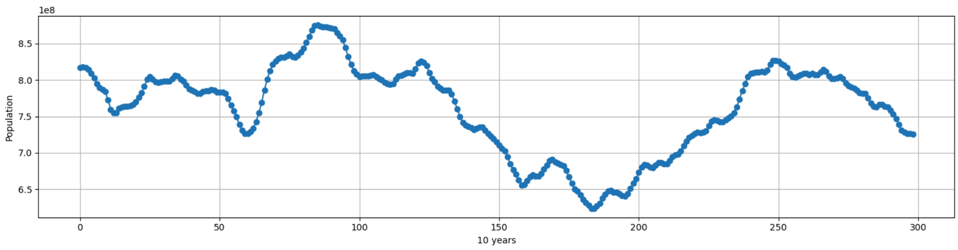


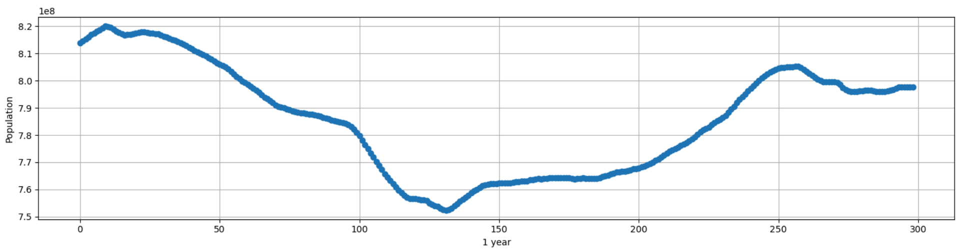
[3]本文所有资料均上传至 https://github.com/1kri2kri/chaos-fractal











可以发现，在时间尺度较长（≥ 1000 years）时，人口数变化随机、完全无法预测，出现混沌现象。而时间尺度较短（≤ 10 years）时，人口数变化则较为连续。这样的结果符合预期。

在这样的简单模型下，预测长时间后的人口都是完全不可行的。由此可见对人口数等历史变量的建模之难。

五　总结

本文通过研究人口这个指标以揭示混沌理论与历史研究间的关系。

文章先介绍并分析了 Logistic 函数这个经典模型，及其离散情况下的变种，再用这个模型（以及其改进版）去拟合人口分布曲线，最后，通过虚拟的"历史游戏"展示出影响因素较多时，长时间尺度下人口的混沌现象。从这些侧面，反映了混沌理论与历史的关系。

历史的发展是受到各种因素影响的，地球上 80 亿人共同引导着历史的发展。除了人类，历史也受到各种环境因素的影响制约。这么看来，对历史的定量研究显得困难重重。

但如果不把眼光放在千年之后，而只关心较近的未来的发展情况，那么一些数学模型就能成功起到预测作用。这样的预测，也能真切影响到国家发展战略与人类未来。所以，对其的描述就尤为重要。而对这样复杂事物的描述，一定不会是简单线性的，而要用到非线性的理论。

这就是将混沌、非线性理论与历史研究结合的意义所在。

六　致谢

感谢梁福明老师的悉心教导！

感谢北京大学为我提供学习交流的平台！

参考文献

1. Verhulst, P. F. Notice sur la loi que la population poursuit dans son accroissement. (1838).
2. Bacaër, N. *Verhulst and the logistic equation (1838)*. 35–39 (Springer London, London, 2011). doi:10.1007/978-0-85729-115-8_6.