

如需查看该出版物的讨论、统计资料和作者简介，请访问：<https://www.researchgate.net/publication/236941795>

时间序列预测的机器学习策略

商业信息处理》讲义 *中的章节* - 2013 年 1 月

DOI: 10.1007/978-3-642-36318-4_3

引文

517

阅读

79,793

3 位作者：



吉安卢卡·邦坦皮

布鲁塞尔自由大学

384 篇著作 19 331 次引用

查看简介



Souhaib Ben Taieb

穆罕默德-本-扎耶德人工智能大学

37 篇著作 3,167 次引用

查看简介



Yann-Aël Le Borgne

布鲁塞尔自由大学

80 篇著作 2,568 次引用

查看简介

本页下面的所有内容由 [Gianluca Bontempi](#) 于 2014 年 6 月 04 日上传。

用户要求增强下载的文件。

时间序列预测的机器学习策略

Gianluca Bontempi、Souhaib Ben Taieb 和 Yann-Aël Le Borgne

机器学习小组

布鲁塞尔自由大学 ULB 科学院计算机科学系

比利时布鲁塞尔凯旋门大街 1050 号

{gbonte,sbentaieb,yleborgn}@ulb.ac.be

<http://mlg.ulb.ac.be>

摘要随着大量历史数据的日益增多，以及在多个科学和应用领域对未来行为进行准确预测的需求，需要定义能够从观测数据中推断过去与未来之间随机关系的稳健而高效的技术。自 20 世纪 60 年代以来，预测领域一直受到 ARIMA 模型等线性统计方法的影响。最近，机器学习模型引起了人们的关注，并在预测领域成为经典统计模型的有力竞争者。本章从三个方面概述了时间序列预测中的机器学习技术：将一步预测问题形式化为 supervised learning tasks；讨论局部学习技术作为处理时间数据的有效工具；以及当我们从一步预测转向多步预测时预测策略的作用。

关键词时间序列预测、机器学习、局部学习、懒惰学习、多输入多输出（MIMO）。

1 引言

*时间序列*是以相等的时间间隔对可观测变量 y 进行历史测量的序列 S_t 。研究时间序列有几个目的，例如根据过去的知识预测未来，了解测量结果背后的现象，或者只是简明扼要地描述序列的突出特征。在本章中，我们将只讨论预测问题。预测观测时间序列的未来值在经济、金融、商业智能、气象和电信等几乎所有科学和工程领域都发挥着重要作用[43]。预测任务的一个重要方面体现在预测范围的大小上。如果说对时间序列进行一步预测已经是一项艰巨的任务，那么进行多步预测则更加困难[53]，因为

还会产生更多并发症，如误差累积、精度降低和不确定性增加等 [58,49]。

长期以来，预测领域一直受到 ARIMA 模型等线性统计方法的影响。然而，在 20 世纪 70 年代末和 80 年代初，人们越来越清楚地认识到，线性模型并不适合许多实际应用[25]。在同一时期，人们提出了一些有用的非线性时间序列模型，如双线性模型[44]、阈值自回归模型[56,54,55]和自回归条件异方差（ARCH）模型[22]（综述见[25]和[26]）。然而，与线性时间序列相比，非线性时间序列分析和预测的分析研究仍处于起步阶段[25]。

在过去的二十年里，机器学习模型引起了人们的关注，并在预测界成为经典统计模型的有力竞争者[1,43,61]。这些模型也被称为黑箱模型或数据驱动模型[40]，是非参数非线性模型的范例，它们仅使用历史数据来学习过去与未来之间的随机依赖关系。例如，Werbos 发现人工神经网络（ANN）优于线性回归和盒式詹金斯方法等经典统计方法[59,60]。Lapedes 和 Farber 也进行了类似的研究[33]，他们得出的结论是，人工神经网络可成功用于非线性时间序列的建模和预测。后来又出现了决策树、支持向量机和近邻回归等其他模型[29,3]。此外，一些机器学习模型的经验准确性已在不同数据条件下的预测竞赛（如 NN3、NN5 和年度 ESTSP 竞赛 [19,20,34,35]）中进行了探讨，在数据挖掘和预测领域引发了有趣的科学辩论 [28,45,21]。

本章旨在概述机器学习技术在时间序列预测中的作用，重点关注三个方面：将一步预测问题形式化为监督学习任务，讨论局部学习技术作为处理时间数据的有效工具，以及当我们从一步预测转向多步预测时预测策略的作用。

本章大纲如下。第 2 节介绍了时间序列建模的一些基本概念，以及将预测任务形式化为输入-输出问题。第 3 节讨论了机器学习技术在从观测数据中推断准确预测结果方面的作用，并介绍了局部学习范式。第 4 节介绍了迄今为止文献中提出的几种多步骤预测策略。第 5 节回顾了如何将局部学习技术与多步骤策略相结合，以进行精确的多步骤预测。

2 预测和建模

基于历史数据集的预测问题主要有两种解释。统计预测理论假定观测到的序列

是随机过程的具体实现，其中的随机性来自于许多独立自由度的线性相互作用[4]。然而，动力系统理论[23,17]中出现的新观点认为，表面上看似随机的行为可能是由只有少量自由度、非线性相互作用的确定性系统产生的。这种复杂的非周期性行为也被称为**确定性混沌**[48]。

我们的工作假设是，许多类别的实验时间序列都可以在动态系统方法的框架内进行分析。因此，时间序列被解释为一个动态系统的观测值，该系统的状态 s 在状态空间 $\Gamma \subset \mathbb{R}^g$ 中演化，其规律为

$$s(t) = F^t(s(0)) \quad (1)$$

其中， $F : \Gamma \rightarrow \Gamma$ 是表示动力学的映射， F^t 是其迭代版本， $s(t) \in \Gamma$ 表示时间 t 的状态值。

在没有噪音的情况下，时间序列与动力系统的 关系为

$$y_t = G(s(t)) \quad (2)$$

其中 $G : \Gamma \rightarrow \mathbb{R}^D$ 称为**测量函数**， D 是序列的维数。下面我们只讨论 $D = 1$ 的情况（**单变量时间序列**）。

函数 F 和 G 都是未知的，因此一般来说，我们无法指望重建原始状态。不过，我们或许可以重新创建一个在某种意义上等同于原始状态的状态空间。

状态空间重构问题包括在观测序列 y_t 中包含唯一可用信息时重构状态。状态空间重构由 Packard 等人[42] 和 Takens [52]独立引入动力系统理论。塔肯斯定理意味着，对于一大类确定性系统，存在一个映射（**延迟重构映射**） $\Phi : \Gamma \rightarrow \mathbb{R}^n$

$$\Phi(s(t)) = \{G(F^{-d}(s(t))), \dots, G(F^{-d-n+1}(s(t)))\} = \{y_{t-d}, \dots, y_{t-d-n+1}\} \quad (3)$$

时间序列的一个有限窗口 $\{y_{t-d}, \dots, y_{t-d-n+1}\}$ （**嵌入向量**）与该序列的动态系统状态之间的有限窗口，其中 d 称为**滞后时间**， n （**阶数**）是考虑的过去值的数量。塔肯斯（Takens）指出，当 $n \geq 2g + 1$ 时， Φ 通常是一个**嵌入**，其中嵌入只适用于具有光滑逆的光滑——微分映射[17]。其主要结果是，如果 Φ 是一个嵌入，那么在重构向量空间中会诱导出光滑动力学 $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$

$$y_t = f(y_{t-d}, y_{t-d-1}, \dots, y_{t-d-n+1}) \quad (4)$$

这意味着重建后的状态可用来估算 f ，而 f 则可以替代 F 和 G ，用于时间序列分析、定性描述、预测等任何用途。

(4) 表示法不考虑任何噪声成分，因为它假定确定性过程 f 可以准确描述时间序列。但请注意，这只是表示时间序列现象的一种可能方式，不应先验地否定任何其他表示方法。事实上，一旦我们假设无法获得函数 f 的精确模型，就完全可以将在确定性公式 (4) 扩展为统计非线性自回归 (NAR) 公式

$$y_t = f(y_{t-d}, y_{t-d-1}, \dots, y_{t-d-n+1}) + w(t) \quad (5)$$

在本章接下来的内容中，我们将把公式 (5) 作为时间序列的一般表示方法，其中也包括 (4) 的特殊情况。

从一组观测数据出发的重建方法能否成功，取决于近似 f 的假设、阶次 n 和滞后时间 d 的选择。

在下面的章节中，我们将只讨论 f 的建模问题，假定 n 和 d 的值是先验的。Casdagli 等人的论文[17]提供了关于阶次选择的很好参考。

3 建立时间依赖关系模型的机器学习方法

3.1 监督学习设置

(5) 中的嵌入式表述表明，一旦有了历史记录 S ，一步预测问题就可以作为监督学习问题来解决。监督学习是在有限观测数据的基础上，对一组输入变量和一个或多个输出变量之间的关系进行建模，而输出变量在一定程度上依赖于输入变量。一旦有了映射 (5) 模型，就可以用于一步预测。在单步预测中，序列的 n 个前值，预测问题可以采用一般回归问题的形式，如图 1 所示。

对输入/输出现象（标量输出和矢量输入）进行建模的一般方法依赖于一组观测对（通常称为训练集）的可用性。

在预测设置中，训练集由历史序列 S 通过创建 $[(N - n - 1) \times n]$ 输入数据矩阵

$$X = \begin{bmatrix} y_{N-1} & y_{N-2} & \dots & y_{N-n+1} \\ y_{N-2} & y_{N-3} & \dots & y_{N-n-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_1 & y_n & \dots & y_1 \end{bmatrix} \quad (6)$$

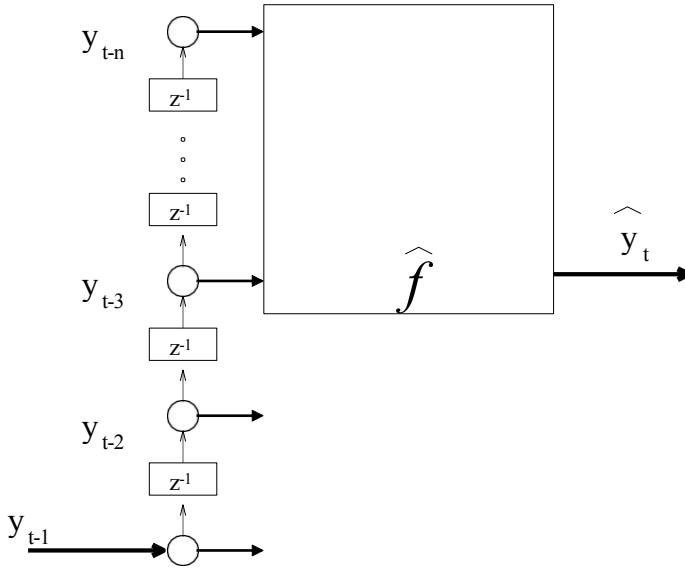


图 1.一步预测 ing。近似器 \hat{f} 返回 预测时间序列在 $t + 1$ 时的值与之前 n 个值的函数关系（包含 z 的矩形框 $^{-1}$ 代表单位延迟算子，即 $y_{t-1} = z^{-1} y_t$ ）。

和 $[(N - n - 1) \times 1]$ 输出向量

$$Y = \begin{bmatrix} y_N \\ y_{N-1} \\ \vdots \\ y_{n+1} \end{bmatrix} \quad (7)$$

为简单起见，我们假设滞后时间为 $d = 0$ 。因此，在本章中，我们将把 X 的第 i^{th} 行（本质上是序列的时间模式）称为 $t - i + 1$ 时间序列的（重建）状态。

3.2 本地学习的实例化

因此，"提前一步预测 "就是在给出过去观测值的子集（也称为查询）时预测输出值。机器学习提供了一个理论框架，可以从观测数据中估算出时间依赖性 f 的合适模型。由于无法在此回顾机器学习在时间序列预测中的全部先进技术，我们将在下一节中更具体地考虑局部学习技术 [12,31,29]。选择这种方法的原因如下：

- 减少假设：本地学习不需要先验地了解数据的基本过程。例如，它不

假设存在一个描述数据的全局函数，不假设噪声的属性。唯一可用的信息是一组有限的输入/输出观测数据。这一特点在实际数据集中尤为重要，因为在实际数据集中，特征缺失、非平稳性和测量误差等问题使得数据驱动和无假设的方法显得尤为重要。

- 在线学习能力：局部学习法可以轻松应对训练样本数量随时间增加的在线学习任务。在这种情况下，局部学习只需向数据集添加新的点，而无需在有新数据时进行耗时的重新训练。
- 非平稳性建模：局部学习法可以处理时变配置，即数据的基本随机过程是非稳态的。在这种情况下，只需从空间和时间两个角度来解释“邻居”的概念即可。对于每个查询点来说，邻近样本不再是具有相似输入的样本，而是具有相似输入且在时间上最近收集到的样本。因此，时间变量成为准确预测需要考虑的另一个宝贵特征。

下面我们将介绍两种局部学习技术，即近邻学习技术 [36,29] 和懒惰学习技术 [12,5]。

最近邻法近邻法是局部逼近法应用于时间序列预测问题的最普通的例子。这种方法是在数据集中寻找与当前状态最近的邻居，并预测当前状态将以与邻居相同的方式发展。

图 2 是近邻一步预测的一个例子。假设我们有一个截至时间 $t-1$ 的时间序列 y_t ，我们想预测该序列的下一个值。一旦选定了某一维度 n ，例如 $n=6$ ，近邻方法就会搜索过去的模式，即

在给定的度量中，与 $\{y_{t-6}, y_{t-5}, \dots, y_{t-1}\}$ （虚线）最相似。例如，如果最近的模式是 $\{y_{t-16}, y_{t-15}, \dots, y_{t-11}\}$ ，那么 NN 方法返回的预测值 \hat{y}_t 就是 y_{t-10} （黑点）。

这种方法最早由 Lorenz [36] 提出，用于研究天气图。

假设我们要预测布鲁塞尔明天的天气，我们选择的维度为 $n=1$ 。近邻法建议：(i) 搜索布鲁塞尔气象条件的历史数据库；(ii) 找到与今天最相似的天气模式（例如 1999 年 3 月 5 日的天气模式，碰巧是个雨天！）；(iii) 预测明天的天气将与 1999 年 3 月 6 日相同（碰巧又是个雨天！）。

近邻法的自然扩展考虑了更多的近邻[31]或更高阶的近似。Tong 和 Lim [56] 在时间序列分析中引入了分段线性近似。Priestley [46] 提出了高阶近

似的重要性。Farmer 和 Sidorowich[23,24]。

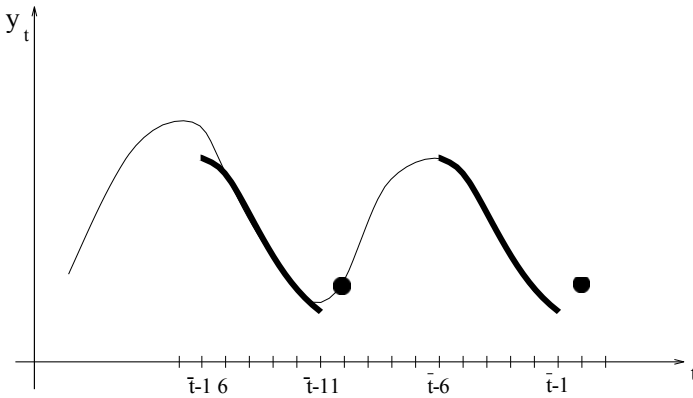


图 2.近邻一步前瞻预测。我们要预测 $t-1$ 时 $n=6$ 阶序列 y 的下一个值。 $y_{t-16}, y_{t-15}, \dots, y_{t-11}$ 与 $\{y_{t-6}, y_{t-5}, \dots, y_{t-1}\}$ 模式最相似。然后，返回预测结果 $\hat{y}_{t-1} = y_{t-10}$ 。

研究了时间序列中的局部逼近，并在一些实验和数值时间序列分析中证明了其有效性。特别是，他们应用局部学习技术来预测混沌时间序列的行为，这些序列虽然是确定的，但其特征是具有随机特性的二阶每秒统计。

懒惰学习懒惰学习（LL）是一种懒惰的局部学习机[12,11]，它会根据交叉验证标准自动调整邻域的大小。懒惰学习的主要魅力在于它的分而治之特性：懒惰学习将复杂的非线性建模问题简化为一系列易于管理的局部线性问题，每个查询都有一个问题。这样就可以在局部基础上利用各种线性识别和验证技术，这些技术快速、可靠，并具有丰富的理论分析、理由和保证。一旦给出 (6) 和 (7) 中的矩阵 X 以及查询点 x_q ，懒惰学习过程主要包括以下步骤：

1. 根据 x 与 q 之间的距离（如欧氏距离），对 X 中的向量集合进行排序。
2. 确定最佳邻居数量。
3. 在给定邻居数量的情况下，使用本地模型（如常数或线性模型）计算查询点的预测值。

让我们考虑一个时间序列 $\{y_1, \dots, y_t\}$ 由 t 个观测值组成，我们打算预测下一个观测值。

预测问题可归结为，当最新的观测窗口由向量 $\mathbf{x}_q = \{y_t, \dots$ 时，估计输出 $y_{t+1} \circ, y_{t-n+1}\}$ 。算法 1 演示了恒定局部学习技术如何返回输出结果。

与查询点 \mathbf{x} 相关的 q ，对于给定的邻居数量 k 。

$[j]$ 用来表示 \mathbf{x} 的第 j 个近邻的索引 q 。请注意，正如 [11] 中所讨论的，该算法的局部线性版本也很常用。

算法 1.LL
<p>输入: $D = \{(\mathbf{x}_i, y_i) \in (\mathbb{R}^n \times \mathbb{R})\}$，数据集。</p> <p>输入: $\mathbf{x}_q \in \mathbb{R}^d$，查询点。</p> <p>输入: k = 邻居数量。</p> <p>输出: $y_{\cdot t+1}$，查询点 \mathbf{x}_q 输出的估计值 (由 k 个邻居获得)。</p> <p>根据与 $\sum_{j=1}^k$ 的距离对向量集合 $\{\mathbf{x}_i\}$ 进行排序。</p> <p>返回 $y_{\cdot t+1}$。</p>

这种算法需要选择一组模型参数（如邻居数 k 、核函数、距离度量）[5]。在此，我们将讨论一种基于“留出一个”（LOO）准则的自动方法来确定邻居数量[11,12]。其主要思路是使用 LOO 指标评估每个局部模型的质量，并根据该指标选择最佳邻域大小。

艾伦（Allen）[2]于 1974 年提出的 PRESS 统计量是一种高效的计算方法，可用于执行 LOO 交叉验证和评估局部线性模型的泛化性能。通过评估每个局部模型的性能，可以对其他配置进行测试和比较，从而选出预期预测效果最好的一个。其原理是将 LOO 误差 $e_{LOO}(k)$ 与估计值相关联。

$$\hat{y}_q^{(k)} = \frac{1}{k} \sum_{j=1}^k y_{[j]} \quad (8)$$

与查询点 \mathbf{x} 相关联 q ，并由 k 个邻居返回。在恒定模型的情况下，LOO 项的推导过程如下[12]：

$$e_{LOO}^{(k)} = \frac{1}{k} \sum_{j=1}^k (e_j^{(k)})^2 \quad (9)$$

其中

$$e_j^{(k)} = y_{[j]} - \frac{\sum_{i=1(i \neq j)}^k y_{[i]}}{k-1} = k \frac{y_{[j]} - \hat{y}_k}{k-1} \quad (10)$$

最佳邻居数定义为

$$k^* = \arg \min_{k \in \{2, \dots, K\}} e_{LOO}(k), \quad (11)$$

使 LOO 误差最小。

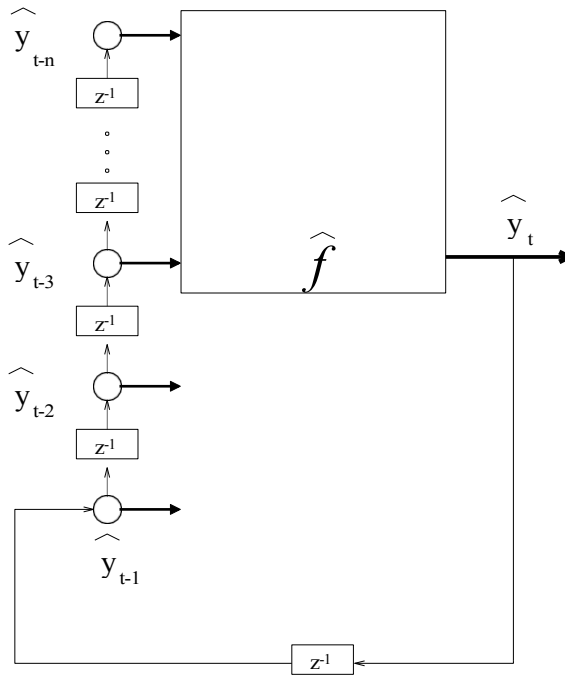


图 3.迭代预测。近似器 \hat{f} 通过迭代前几步得到的预测值，返回时间 $t + 1$ 时时间序列值的预测值（包含 z^{-1} 的矩形 box 代表单位延迟算子，即 $\hat{y}^{t+1} = z^{-1} \hat{y}^t$ ）。

懒惰学习被成功应用于多项回归和一步预测任务 [14]。有关 LL 技术及其应用的更多详情，请参阅 [11,12]。

4 多步骤时间序列预测策略

上一节表明，一步预测可以通过求助于常规学习技术（如局部学习）在常规监督学习框架内进行。在本节中，我们将扩展这一框架，说明如何利用学习技术来解决多步骤预测问题。我们可以考虑三种策略，即递归策略、直接策略和多重输出策略。

多步骤时间序列预测任务包括预测由 N 个观测值组成的历史时间序列 $[y_1, \dots, y_N]$ 的下 H 个值 $[y_{N+1}, \dots, y_{N+H}]$ ，其中 $H > 1$ 表示预测范围。

本节将介绍在多步骤预测中采用机器学习的三种现有策略。我们将使用一个通用的符号，其中 f 和 F 表示过去和未来观测值之间的函数依赖关系、

n 是指时间序列的 embedding 维度[17]，即用于预测未来值的过去值的数量，而 w 则代表包含建模误差、干扰和/或噪声的项。

4.1 递归策略

递归策略[58,49,18] 首先训练一步模型 f

$$y_{t+1} = f(y_t, \dots, y_{t-n+1}) + w_{t+1} \quad (12)$$

$t \in \{n, \dots, N-1\}$ ，然后使用它递归返回多步预测结果（图 3）。递归法的一个众所周知的缺点是对估计误差的敏感性，因为当我们在未来更进一步时，会更多地使用估计值，而不是实际值 d 。

尽管存在这些局限性，递归策略仍被成功地用于预测现实世界中的许多时间序列，使用了不同的机器学习模型，如递归神经网络 [47] 和最近邻网络 [38,15]。

4.2 直接策略

直接策略[58,49,18] 独立学习 H 模型 f_h

$$y_{t+h} = f_h(y_t, \dots, y_{t-n+1}) + w_{t+h} \quad (13)$$

$t \in \{n, \dots, N-H\}$ ， $h \in \{1, \dots, H\}$ ，并通过连接 H 个预测返回多步预测。

由于直接策略不使用任何近似值来计算预测值（等式 13），因此不容易积累误差。尽管如此，它也有一些弱点。首先，由于 H 模型是独立学习的，因此没有考虑预测值 \hat{y}_{N+h} [13,16,32] 之间的统计依赖关系。其次，直接方法往往需要比 ted 方法更高的函数复杂度[54]，以模拟两个较远时刻的两个序列值之间的随机依赖关系[27]。最后但并非最不重要的一点是，这种策略需要大量的计算时间，因为需要学习的模型数量等于水平线的大小。

不同的机器学习模型被用于实施多步骤预测任务的直接策略，例如神经网络 [32]、近邻 [49] 和决策树 [57]。

4.3 DirRec 战略

DirRec 策略[50] 结合了直接策略和递归策略的结构和原理。*DirRec* 在每个时间跨度用不同的模型计算预测（与直接策略类似），在每个时间步，它通过添加与上一步预测相对应的变量来扩大输入集（与递归策略类似）。然而，需要注意的是，与前两种策略不同，嵌入大小 n 并非对所有的

地平线。换句话说, DirRec 策略从时间序列 $[y_1, \dots, y_N]$ 中学习 H 模型 f_h , 其中

$$y_{t+h} = f_h(y_{t+h-1}, \dots, y_{t-n+1}) + w_{t+h} \quad (14)$$

$t \in \{n, \dots, N-H\}$, $h \in \{1, \dots, H\}$ 。

4.4 多种输出策略

多步预测的迭代技术和直接技术尽管各有不同, 但都有一个共同特点: 它们都是从数据中建立多输入单输出映射模型, 迭代技术的输出是变量 y_{t+1} , 直接技术的输出是变量 y_{t+k} 。当需要进行长期预测并假定存在随机环境时, 单输出映射建模会忽略未来值之间 (如 y_{t+k} 和 y_{t+k+1} 之间) 存在的随机依赖关系, 从而使预测精度出现偏差。弥补这一缺陷的可行方法是从单输出映射建模转向多输出依赖性建模。这就需要采用多输出技术, 即预测值不再是标量, 而是时间序列未来值的向量。

多输入多输出 (MIMO) 策略。 多输入多输出 (MIMO) 策略[13,16] (也称为联合策略[32]) 通过学习单一的多输出模型, 避免了直接策略[13,16]对未来值之间条件独立性的简单假设

$$[y_{t+H}, \dots, y_{t+1}] = F(y_t, \dots, y_{t-n+1}) + w_t \quad (15)$$

其中, $t \in \{n, \dots, N-H\}$, $F: \mathbb{R}^d \rightarrow \mathbb{R}^H$ 是一个向量值函数[39], 并且 $w_t \in \mathbb{R}^H$ 是一个噪声矢量, 其协方差不一定是对角的[37]。

预测结果由多输出模型 \hat{F} 一步返回, 其中

$$[\hat{y}_{t+H}, \dots, \hat{y}_{t+1}] = \hat{F}(y_N, \dots, y_{N-n+1}). \quad (16)$$

MIMO 策略的基本原理是在预测值之间模拟时间序列的随机依赖性。这种策略避免了直接策略中的条件独立性假设, 也避免了递归策略中的误差累积。迄今为止, 这一策略已成功应用于现实世界中的多项多步骤时间序列预测任务 [13,16,10,9]。

然而, 由于希望保留随机依赖关系, 因此所有水平期的预测都必须采用相同的模型结构。由于这种限制可能会降低预测方法的灵活性[10], 因此下文将讨论 MIMO 策略的一种变体。

DIRMO 策略。DIRMO 策略[10,9]旨在保留 DIRect 和 MIMO 策略最吸引人的地方，它将视界 H 划分为几个区块，并使用 MIMO 预测每个区块内的值。

块。这意味着 H 步预测需要 m 个多输出预测任务 ($m = H$)，每个任务的输出大小为 s ($s \in \{1, \dots, H\}$)。

请注意，当 $s = 1$ 时，DIRMO 与传统的直接策略相吻合，而当 $s = H$ 时，则与 MIMO 策略相对应。通过调整参数 s ，我们可以校准输出维度 ($s = 1$ 时无依赖性， $s = H$ 时依赖性最大)，从而提高 MIMO 策略的灵活性。这样，既能保留未来值之间较大程度的随机依赖性，又能提高预测器的灵活性，可谓一举两得。

5 多步骤预测的本地学习

局部学习似乎不仅是单步预测的有效算法，也是多步预测的有效算法。本节将讨论一些使用局部学习技术专门处理长期预测问题的著作。

在文献[38,15]中，作者提出了对局部学习技术的修改，以考虑多步骤预测问题的时间行为，从而改进递归策略的结果。特别是

[15] 通过引入迭代版本的留一统计量，对 PRESS 准则 (10) 进行了修改。他们的研究表明，通过评估局部一步预测器在长于一步的范围内的泛化性能，迭代 PRESS 优于非迭代标准，同时还保留了计算效率的良好特性。值得注意的是，[38]和[15]提出的两种技术在 1998 年鲁汶时间序列预测中分别排名第一和第二。

RECNOISY [6]是对基于局部学习的递归策略的最新改进，它在预测过程的每一步都对初始数据集进行扰动，以更恰当地处理预测过程中的近似值。RECNOISY 方法的基本原理是，递归策略所使用的训练示例虽然是观察到的，但并不一定能代表以后整个预测过程中所需的预测任务。为了解决这一问题，该策略利用了递归策略引起的预测任务的特殊性，并将其纳入局部学习阶段，以改善结果。

文献[51]介绍了懒惰学习的两种改进方法，用于处理时间序列的长期预测。第一种方法基于对输入的迭代修剪；第二种方法是使用 k -NN 近似器在可能的输入集合中进行蛮力搜索。

文献[13]提出了在多输入多输出预测中使用局部学习的方法，并讨论了算法 1 的多输出扩展以及多个长期预测因子的平均策略，以提高预测结果的

准确性。

文献[8,9]介绍了如何利用局部学习近似器实施 DIRMO 策略。基于局部学习的 DIRMO 策略已成功应用于两次预测竞赛：ESTSP'07 [10] 和 NN3 [9]。

并比较了基于局部学习算法的多步骤时间序列预测策略。

6 结论

预测未来是应用科学中最具相关性和挑战性的任务之一。要从历史数据中建立有效的预测工具，就必须采用推断观察值的过去值和短期未来值之间的依赖关系的组合方法和统计方法，以及处理更长远问题的适当策略。本章讨论了机器学习在调整监督学习技术以处理预测问题方面的作用。我们特别强调了局部学习近似值在处理预测中的重要问题（如非线性、非平稳性和误差累积）时所发挥的作用。未来的研究应关注将这些技术扩展到商业智能的一些新方向，如海量数据（大数据）的并行挖掘[41]和应用于空间-时间任务[30]。

nts。 Gianluca Bontempi 感谢比利时法语社区资助的 ARC 项目 "利用功能基因组学和生物信息学发现调控糖尿病胰岛 β 细胞功能障碍和 apoptosis 的分子途径"。

参考资料

1. Ahmed, N.K., Atiya, A.F., El Gayar, N., El-Shishiny, H.: An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews* 29(5-6) (2010)
2. Allen, D.M.: 变量选择与数据分析之间的关系及预测方法。 *Technometrics* 16(1), 125-127 (1974)
3. Alpaydin, E.: 《机器学习导论》，第 2 版。 Adaptive Computation and Machine Learning. 麻省理工学院出版社（2010 年 2 月）
4. Anderson, T.W.: The statistical analysis of time series. J. Wiley and Sons (1971)
5. Atkeson, C.G., Moore, A.W., Schaal, S.: Locally weighted learning. *Air* 11(1-5), 11-73 (1997)
6. Ben Taieb, S., Bontempi, G.: 通过每轮数据进行递归多步时间序列预测。 In: IEEE-ICDM 2011 (2011 年) 论文集
7. Ben Taieb, S., Bontempi, G., Atiya, A., Sorjamaa, A.: 基于 NN5 预测竞赛的多步骤超前时间序列预测策略回顾与比较。 *ArXiv e-prints* (2011 年 8 月)
8. Ben Taieb, S., Bontempi, G., Sorjamaa, A., Lendasse, A.: 通过结合直接策略和 mimo 策略对时间序列进行长期预测。 In: 2009 年电气和电子工程师

学会国际神经网络联合会议论文集》，美国亚特兰大，第 3054-3061 页（2009 年 6 月）

9. Ben Taieb, S., Sorjamaa, A., Bontempi, G.: Multiple-output modelling for multi-step-ahead forecasting. *Neurocomputing* 73, 1950-1957 (2010)

10. Ben Taieb, S., Bontempi, G., Sorjamaa, A., Lendasse, A.: 通过结合直接和 mimo 策略对时间序列进行长期预测。In: 神经网络国际联合会议 (2009 年)
11. Birattari, M., Bontempi, G., Bersini, H.: 懒惰学习与递归最小二乘法算法。In: Kearns, M.S., Solla, S.A., Cohn, D.A. (eds.) NIPS 11, pp.麻省理工学院 出版社, 剑桥 (1999 年)
12. Bontempi, G.: Local Learning Techniques for Modeling, Prediction and Control. 博士论文, IRIDIA-布鲁塞尔自由大学 (1999 年)
13. Bontempi, G.: 利用多输入多输出局部学习进行长期时间序列预测。In: 第二届欧洲时间序列预测 (TSP) 研讨会论文集, ESTSP 2008, 芬兰赫尔辛基, 第 145-154 页 (2008 年 2 月)
14. Bontempi, G., Birattari, M., Bersini, H.: 工作中的懒惰学习者: 懒惰学习工具箱。In: 第七届欧洲智能技术和软计算大会论文集, EUFIT 1999 (1999 年)
15. Bontempi, G., Birattari, M., Bersini, H.: 用于迭代时间序列预测的局部学习。In: Bratko, I., Dzeroski, S. (eds.) Machine Learning: 第十六届国际会议论文集, 第 32-38 页。摩根考夫曼出版社, 旧金山 (1999)
16. Bontempi, G., Ben Taieb, S.: 局部学习中多步超前预测的条件依赖策略。国际预测期刊》(2011 年) (出版中, 修正证明)
17. Casdagli, M., Eubank, S., Farmer, J.D., Gibson, J.: State Space reconstruction in the presence of noise.PHYD 51, 52-98 (1991)
18. Cheng, H., Tan, P.-N., Gao, J., Scripps, J.: Multistep-Ahead Time Series Prediction.In: Ng, W.-K., Kitsuregawa, M., Li, J., Chang, K. (eds.) PAKDD 2006.LNCS (LNAI), vol. 3918, pp.Springer, Heidelberg (2006)
19. Crone, S.F.: NN3 预测竞赛、
<http://www.neural-forecasting-competition.com/NN3/index.html> (更新日期: 2009 年 5 月 26 日) (访问日期: 2010 年 7 月 05 日)
20. Crone, S.F.: NN5 预测竞赛、
<http://www.neural-forecasting-competition.com/NN5/index.html> (更新日期: 2009 年 5 月 27 日) (访问日期: 2010 年 7 月 05 日)
21. Crone, S.F.: 挖掘过去, 决定未来: 评论。International Journal of Forecasting 5(3), 456-460 (2009); Special Section: 时间序列监测
22. Engle, R.F.: Autoregressive conditional heteroscedasticity with the estimates of the variance of the United Kingdom inflation.Econometrica 50(4), 987-1007 (1982)
23. Farmer, J.D., Sidorowich, J.J.: 预测混沌时间序列。物理评论快报》8(59), 845-848 (1987)
24. Farmer, J.D., Sidorowich, J.J.: Exploiting chaos to predict the future and reduce noise.技术报告, 洛斯阿拉莫斯国家实验室 (1988 年)

25. De Gooijer, J.G., Hyndman, R.J.: 25 years of time series forecasting.《国际预测期刊》22 (3) , 443-473 (2006 年)
26. De Gooijer, J.G., Kumar, K.: Some recent developments in non-linear time series modelling, testing, and forecasting.《International Journal of Forecasting》8(2), 135-156 (1992)
27. Guo, M., Bai, Z., An, H.Z.: 基于经验分布的非线性自回归模态多步预测。In. Statistica Sinica, pp: Statistica Sinica, pp.
28. Hand, D.: Mining the past to determine the future: 问题与可能性。《国际预测期刊》(2008 年 10 月)

29. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning: data mining, inference and prediction, 2nd edn. Springer (2009). Springer (2009)
30. Hsu, W., Lee, M.L., Wang, J.: Temporal and spatio-temporal data mining. IGI 出版 (2008 年)
31. Ikeguchi, T., Aihara, K.: Prediction of chaotic time series with noise. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E78-A(10) (1995)
32. Kline, D.M.: 用神经网络进行多步骤时间序列预测的方法。
In: Peter Zhang, G. (ed.) Neural Networks in Business Forecasting, pp. Information Science Publishing (2004)
33. Lapedes, A., Farber, R.: 使用神经网络的非线性信号处理: 预测、预测和预测。
技术报告 LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM (1987). 技术报告 LA-UR-87-2662, 洛斯阿拉莫斯国家实验室, Los Alamos, NM (1987)
34. Lendasse, A. (ed.): ESTSP 2007: 论文集 (2007 年)
35. Lendasse, A. (ed.): ESTSP 2008: Proceedings. Multiprint Oy/Otamedia (2008) ISBN: 978-951-22-9544-9
36. Lorenz, E.N.: 《自然发生的大气变化所揭示的大气可预测性》 (Atmospheric predictability as revealed by naturally occurring analogues. 大气科学杂志, 26, 636-646 (1969 年))
37. Matias, J.M.: 多输出非参数回归。见 Bento, C., Cardoso, A., Dias, G. (eds.) EPIA 2005. LNCS (LNAI), vol. 3808, pp. Springer, 海德堡 (2005 年)
38. McNames, J.: 用于时间序列预测的最近轨迹策略。In: Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling, pp. 比利时鲁汶大学 (1998 年)
39. Micchelli, C.A., Pontil, M.A.: On learning vector-valued functions. Neural Comput. 17(1), 177-204 (2005)
40. Mitchell, T.M.: Machine Learning. 麦格劳-希尔, 纽约 (1997 年)
41. Owen, S.: Mahout in action. 曼宁 (2012)
42. Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S.: 从时间序列看几何物理评论快报》 45(9), 712-716 (1980)
43. Palit, A.K., Popovic, D.: 时间序列预测中的计算智能》:
理论与工程应用。工业控制进展》。 Springer-Verlag New York, Inc., Secaucus (2005)
44. Poskitt, D.S., Tremayne, A.R.: The selection and use of linear and bilinear time series models. 国际预测期刊》, 2 (1), 101-114 (1986 年)
45. Price, S.: 挖掘过去, 决定未来: 评论。国际期刊
ing 25(3), 452-455 (2009)
46. Priestley, M.B.: Non-linear and Non-stationary time series analysis. 学术出版社 (1988 年)
47. Saad, E., Prokhorov, D., Wunsch, D.: 股票趋势预测比较研究
使用时延、递归和概率神经网络。电气和电子工程师学会神经网络论文集》

- 9 (6) , 1456-1470 (1998 年)
48. Schuster, H.G.: Deterministic Chaos: 导论》。Weinheim Physik (1988)
49. Sorjamaa, A.、Hao, J.、Reyhani, N.、Ji, Y.、Lendasse, A.: 时间序列长期预测方法。神经计算 70(16-18), 2861-2869 (2007)
50. Sorjamaa, A., Lendasse, A.: 使用 dirrec 策略进行时间序列预测。In: Verleysen, M. (ed.) European Symposium on Artificial Neural Networks, ESANN 2006, Bruges, Belgium, April 26-28, pp.
51. Sorjamaa, A., Lendasse, A., Verleysen, M.: Pruned lazy learning models for time series prediction. In: 欧洲人工神经网络研讨会, ESANN 2005, 第 509-514 页 (2005 年)

52. Takens, F.: Detecting strange attractors in fluid turbulence. In : Dynamical Systems and Turbulence. 柏林施普林格出版社 (1981 年)
53. Tiao, G.C., Tsay, R.S.: 时间序列中非线性和自适应建模的一些进展。《预测杂志》13 (2) , 109-131 (1994 年)
54. Tong, H.: Threshold models in Nonlinear Time Series Analysis. 柏林施普林格出版社 (1983 年)
55. Tong, H.: Non-linear Time Series: 牛津大学出版社 (1990 年)。牛津大学出版社 (1990 年)
56. Tong, H., Lim, K.S.: Threshold autoregression, limit cycles and cyclical data. JRSS B 42, 245-292 (1980)
57. Tran, T.V., Yang, B.-S., Tan, A.C.C.: 使用回归树和神经模糊系统进行机器状态预报的多步超前直接预测。Expert Syst. 36(5), 9378-9387 (2009)
58. Weigend, A.S., Gershenfeld, N.A.: Time Series Prediction: forecasting the future and understanding the past. Addison Wesley, Harlow (1994)
59. Werbos, P.J.: Beyond Regression: 行为科学预测与分析的新工具》。哈佛大学博士论文, 马萨诸塞州剑桥市 (1974 年)
60. Werbos, P.J.: 应用于循环天然气市场模型的反向传播泛化。神经网络 1 (4) , 339-356 (1988 年)
61. Zhang, G., Eddy Patuwo, B., Hu, M.Y.: 人工神经网络预测: 国际预测杂志》14 (1) , 35-62 (1998 年)。国际预测期刊》14 (1) , 35-62 (1998 年)

[查看出版统计](#)