链接地址：<https://doi.org/10.3390/risks7010033>

于文斌 2019310168

深度学习集成Lee-Carter模型

摘要：在死亡率领域，基于Lee-Carter的方法来预测死亡率具有里程碑意义，我们可以定义一个“Lee-Carter模型家族”来介绍该模型的所有发展，包括其第一个公式（1992）比较未来模型性能的基准。在Lee–Carter模型中，kt参数描述随着时间变化的死亡率趋势，对未来死亡率的预测起着重要作用。通常用于kt建模的传统ARIMA过程来描述未来死亡率有明显的局限性。关于预测阶段，为了考虑预计死亡率的非线性，学术界应采取更合理的方法。因此我们提出了一种替代方案：基于深度学习技术的ARIMA过程。更确切地说，为了更准确地把握kt系列随时间变化的模式，我们应用了具有较长短期记忆架构，并集成Lee-Carter模型以提高其预测能力。所提出的方法在预测准确性方面提供了显著的性能，并且还允许为避免时间块的先验选择。的确，删除噪声溢出或数据质量不足的时间。长短期记忆网络的实力关键在于其能够处理这种噪声，并充分复制它的预测趋势能力，因为其自身的架构可以考虑长期的重要因素模式。

关键词：死亡率；深度学习；长短期记忆；Lee-Carter模型；预测

1.简介

死亡率大大促进了人口动态，并且在许多领域至关重要，例如经济，人口统计学和社会科学。自19世纪初以来，发达国家的预期寿命已经显示出持续增长（自第二次世界大战末期以来一直保持稳定的增长速度），其速度通常比预期的要快（Oeppen和Vaupel（2002））。这有给人寿保险公司，养老金计划和涉及福利的社会保障计划，带来了疾病、死亡或生存有关的财务挑战。目前，精算师普遍使用精算技术来预测未来的死亡率。治疗寿命风险以及预测更准确的预期寿命。因为LC模型具有一些优势，例如与其他模型相比参数数量少和鲁棒性强，经常使用它来预测死亡率。该模型的简约性使该模型的死亡率随年龄增长而稳定，在预测阶段防止年龄轮廓线的交叉（有关更多详细信息，请参见Li等人（2009年））。

近年来，由于计算能力的提高，机器学习技术重新兴起。机器学习包含许多从大型数据集中学习的算法功能，对检测未知和无法识别的模式非常有用。尽管增加在不同研究领域的应用，机器学习在人口统计中的应用并不那么受欢迎，主要由于结果难以解释，所以通常被认为是“黑匣子”。此外，机器学习算法只受数据驱动而不受特定假设的支持，人口统计学家对此更感兴趣。

尽管有人口统计学家的储备，但该领域的研究贡献正在增加。我们提醒了Hainaut（2018），他提出了一个神经网络来预测和模拟死亡率。作者使用神经分析仪来识别潜在的时间过程并直接预测死亡率。这种方法可以检测和复制在死亡率的对数力演变过程中观察到的非线性。Deprez等（2017）使用一些机器学习技术来改善对数死亡率的估计，Levantesi和Pizzorusso（2018）将其扩展到Lee-Carter框架中的死亡率预测。此外，Richman和Wüthrich（2018）的最新论文提出了集成Lee-Carter模型，其中使用神经网络估计参数。其中使用神经网络估计参数。精算领域中机器学习的其他相关应用可以在例如Castellani等人的文章中找到。 （2018）和Gabrielli和Wüthrich（2018）。

在本文中，我们使用深度学习技术来提高Lee-Carter模型的预测能力。具体来说，我们的方法旨在整合原始的Lee-Carter公式引入具有长短期记忆（LSTM）架构的递归神经网络进行预测kt参数的未来发展，克服了ARIMA过程显示的局限性。在非线性死亡率趋势的情况下，我们认为使用LSTM可以获取与观察到的死亡率动态更为一致的死亡率预测。更准确地说，LTSM网络的结构是为了详细说明较长的数据序列，从而形成一个能够保留数据之间重要关系的存储器，该关系在序列中也很远。从这个意义上讲，在时间序列范围内，考虑到过去死亡率趋势的重大影响，LSTM可以预测一段时间内的未来死亡率，并将其充分地再现为预测趋势。LSTM的强项就是随着时间的推移保留信息，从而防止较旧的信号在处理过程中逐渐消失。

因此，本文主要关注死亡率趋势的预测，而Lee和Carter（1992）的参数估计保持不变。具体而言，我们没有提出用于拟合死亡率表面的新方法（Hainaut（2018）已经引入，它使用神经网络替代传统奇异值分解来拟合死亡率）;相反，我们引入了基于LSTM网络的创新结构来对未来的死亡率趋势进行建模。

本文脉络如下。第2节介绍了Lee-Carter模型的传统方案，第3节介绍了神经网络，其中特别关注LSTM网络。第4节专门介绍在全世界六个国家和地区中进行的数值应用以及第5节提供结论。

2.LC模型

作者在1992年使用1933-1987年期间的美国死亡率数据开发了Lee-Carter（LC）模型的第一版。在本文中，我们参考模型的第一个版本，然后将奇异值分解（SVD）应用于中心死亡率的对数以获得参数。该模型具有以下表达式：



其中mx，t是在t年中在x年龄观察到的中心死亡率，ax是死亡率，bx是随着kt的变化而与个人年龄的偏离方式，kt是描述死亡率趋势，而ex，t是x岁和t时间的残差项。为了避免参数的可识别性问题，模型需要以下约束：

在原始LC模型中，参数通过奇异值分解（SVD）估算分为两个阶段。 首先，将SVD应用于log（mx，t）􀀀ax的矩阵，以找到bx和kt。 其次，重新安装kt，使观察到的死亡人数与估计的死亡人数一致。在传统的LC公式中，kt通常由ARIMA（0,1,0）建模：

1575511019(1)

其中d是漂移参数，et是误差项，均值和方差均值为零。

3.神经网络模型

神经网络（NN）起源于复制生物学人脑神经网络的数学模型（Minsky和Pitts（1943），Wiener（1948））。 NN架构包括神经元，连接神经元的突触连接和学习算法。 通常NN由三种类型的层分别称为输入层，隐藏层和输出层，每一层都有几个神经元。网络中的每个单元都通过其他连接的突触链接获得“加权”信息并通过使用激活函数转换输入信号的加权总和来返回输出。考虑单个神经元H（Rosenblatt，1958年称为感知器），其输出定义为：



其中x是输入，w是相关突触权重，d 是输入信号的数量f是激活功能。项b表示偏差，也称为激活阈值。重要的是要注意到函数f必须是可微的，因为学习方程是基于梯度的。具有单层的NN不能用于非线性可分离问题，例如XOR（Minsky和Papert（2017）），但多层感知器（MLP）的引入启发了功能更强大的新结构克服了这一限制的神经网络。在MLP中，神经元容易分布在不同的层上单元与上一层完全连接。通过突触连接单元的方式定义了不同类型的网络。在NN中经典模式（前馈NN），信息仅在一个输入层中从输入层移动到输出层方向，而在递归神经网络（RNN）中，使用其他突触来使处理结果作为输入重新处理。

图1为典型的前馈NN表示。 图的每个节点代表一个神经元，通过表示突触的弧线彼此连接。

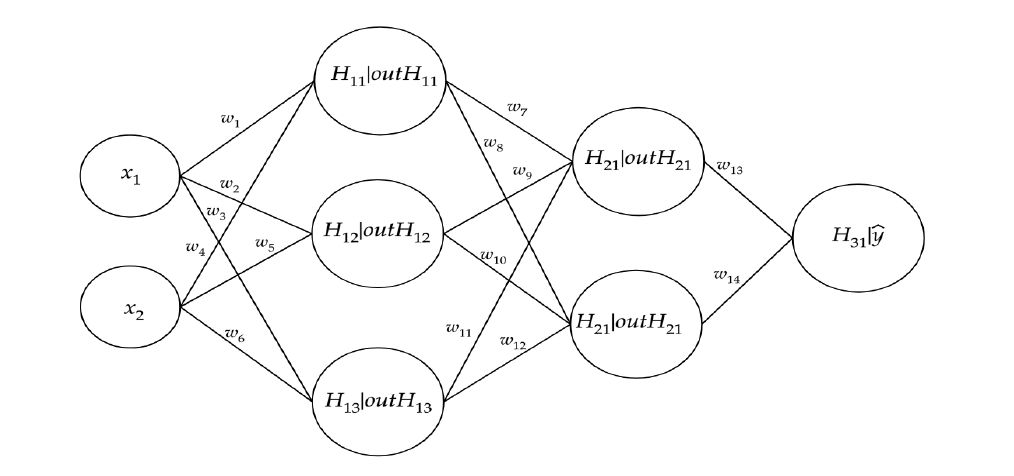


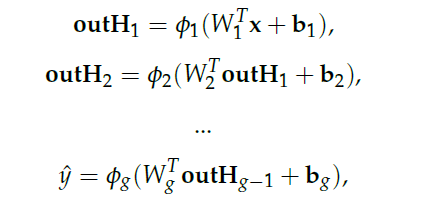
图1.神经网络（NN）的示意图：圆圈代表神经元，线条代表突触。 突触取输入值并乘以“权重”。 神经元添加所有突触的输出并带入激活函数。

具有nh个神经元的隐藏层的输出out定义为：

1575511859(1)

其中W 是权重矩阵，b是偏差。 在MLP方案中隐藏层成为下一层的输入。

考虑一个回归问题，其中g是隐藏层数，输出y为通过以下方式获得：



其中W1，W2，...，Wg是权重矩阵，b1，b2，...，bg是偏差向量，而f1，f2，...，fg是激活功能不一定彼此不同。 请务必注意，权重矩阵和偏差向量取决于隐藏层中的单位数。

反向传播

NN训练涉及优化问题，其目的是使在高维空间中的损失函数最小化。 让我们定义一个损失函数：



它测量了预测值和观测值y之差。数量E取决于在权重1 W1，W2，...，Wg的矩阵上影响y的值。目的是寻找价值最小化数量E的突触权重在许多其他算法中，使用反向传播训练神经网络。该算法将预测值与期望值进行比较（客观）并进行修改通过向后传播损失函数的梯度来传递突触权重。从原理上讲交替向前和向后传播步骤：

在前进步骤中，计算预测值固定突触权重，

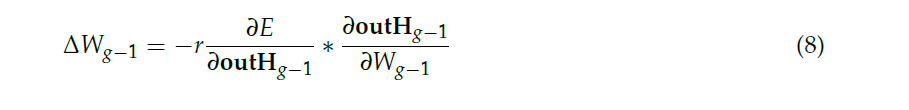
在后退步骤中，调整权重以减少网络的误差E。

NN迭代执行正向和反向传播并修改权重以找到使损失函数最小化的组合。

反向传播算法使用增量更新最后一层的权重Wg规则（Bryson和Ho（1969）），定义为：

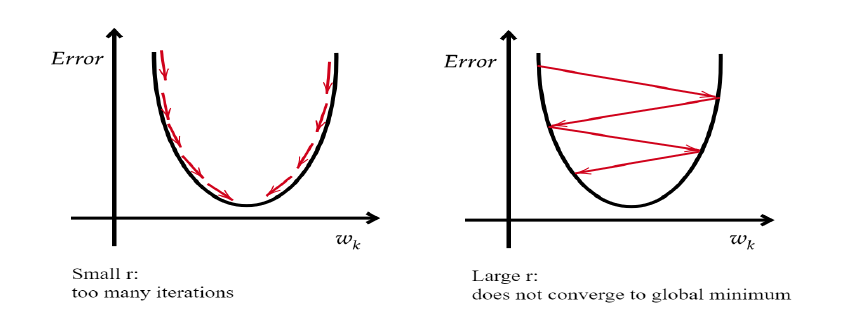


其中r是学习率。 对于其他先前的层，我们继续使用链推导规则。对权重矩阵进行更新：



以此类推，从某种意义上讲梯度下降背后的思想类似于“爬下山坡”，直到达到全局或局部最小值。 每次更新都会移入梯度的相反方向以及梯度的斜率和学习率决定了此运动的幅度。 此外学习率r的选择是一个重要因素，因为值涉及太多的迭代，而较大的值则不允许收敛到全局最小值。

图2显示了最简单情况下学习率在梯度下降收敛中的作用一维空间。



有关架构类型的选择（例如，隐藏层的数量，每层的单位）而超参数值（例如学习率，激活函数，步长）对于NN用户来说仍然是一个问题：选择通常取决于数据类型，这并非易事。在测试之前需要对超参数进行调整。NN和反向传播算法可以在Alpaydin（2010），Hastie等人的文章中找到。

3.1。具有长短期记忆架构的递归神经网络

前馈神经网络虽然代表了强大的分析工具（Hornik et al。（1989）），但它们是不足以有效地管理数据的时间顺序。节点之间的递归连接（以RNN为特征）允许动态顺序数据分析。但是通过使用RNN结构，我们通常会面临梯度消失主要问题，权重发生变化并且变得如此之小以至于没有影响。网络逐渐失去了学习的能力，从而不足以分析长数据序列和做出良好的预测。因此RNN的记忆力很短。为了克服这个问题，Hochreiter和Schmidhuber（1997）开发了所谓的长短期记忆LSTM。 LSTM是一种RNN，其架构使得它可以考虑序列数据之间的关系（即使很长）也消除了消失的梯度问题。通过这种方式RNN既获取了长时记忆，也获取了短时记忆。LSTM经历了几次如（Gers等（1999）和（Gers，Schmidhuber（2000））所示，现在有可能定义了组成良好的基本结构，称为LSTM。根据图3，单个LSTM单元（称为LSTM块）由三个门组成，它们之间的相互作用以及由此产生的存储单元。该块接收当前的信息作为初始信息流输入xt，先前的短期结果ht和（长期）存储单元的先前状态ct。该信息由三个门（红色圆圈）详细说明，分别命名为遗忘门，输入门和输出门以及辅助NN（蓝色圆圈）有助于信息流的规范化。

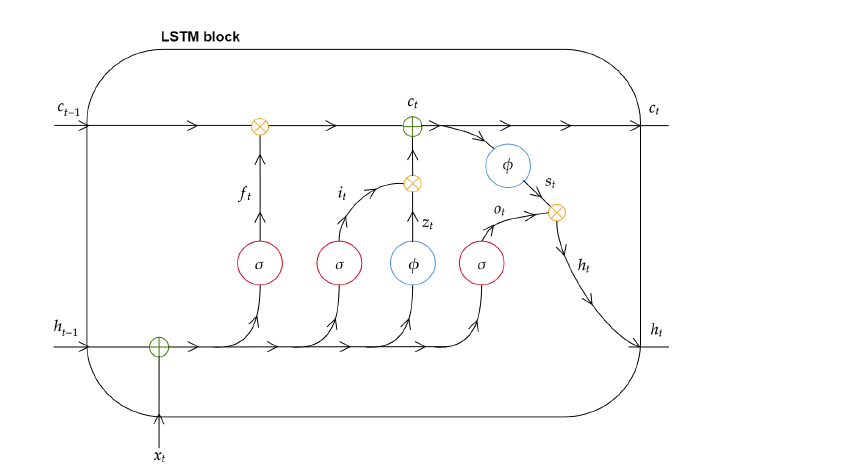
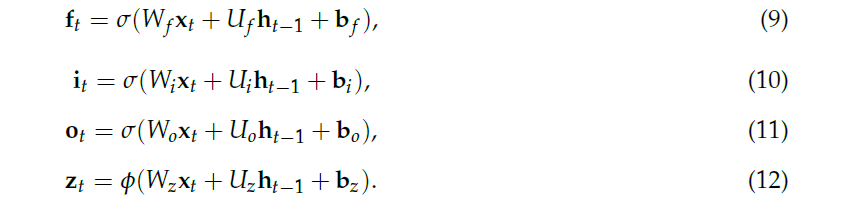
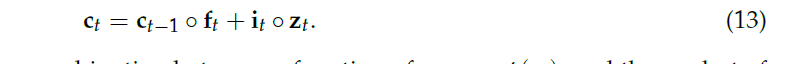


图3.长短期记忆（LSTM）块结构的表示。

形式上，参考由LSTM块的nh∈N组成的隐藏层，令W∈Rd×nh和U∈Rnh×nh权重矩阵（针对每个门）分别与输入和先前的短期结果相关联， LSTM体系结构的递归网络的一般操作模式由以下方程式描述，分别指定了遗忘门输出（ft∈Rnh），输入门输出（it∈Rnh），输出门输出（ot∈Rnh），和辅助输出门的输出（zt∈Rnh）：

由方程式（9）定义的遗忘门输出ft使得来自前一单元状态的信息和来自当前输入的信息通过Sigmoid激活函数以非线性方式混合。因此输出介于0和1之间的值，而忽略或保留前一个块的状态。然后将ft与以前的存储状态ctt 1进行逐点乘积混合。

等式（10）中定义的输入门it也使用sigmoid激活函数，从而可以确定何时应更新接收到的信息。

等式（11）中描述的输出门ot具有防止将不重要的存储器内容存储的信息传输到其他块的作用。 为此使用sigmoid函数以传递相关的存储器信息。为了使处理后的数据流规则化，将输入门与从关联的辅助NN zt获得的输入门组合起来，如式（12）所示。

为了获得当前的输出，需要结合ct的函数st = f（ct）以及与输出门相关的辅助NN是必需的：



输出outH 2 Rnh被传递到下一层，并成为短存储器ht = outH。下一个实例表示每个时间步具有LSTM架构的RNN操作的一般模式在图4中。

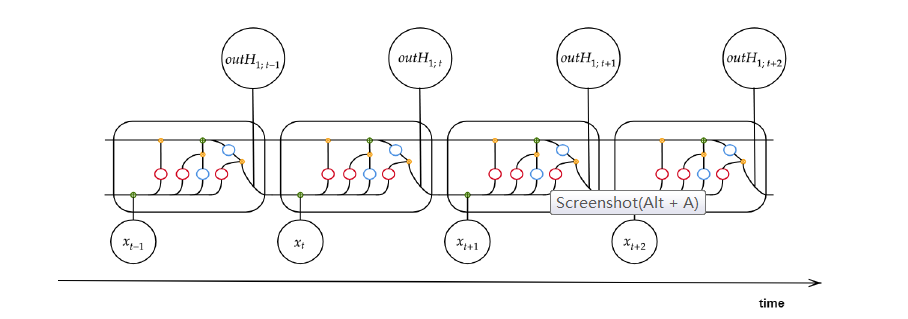


图4.一个隐藏分层递归神经的顺序操作的简化表示每个时间步都有LSTM体系结构的网络（RNN）。

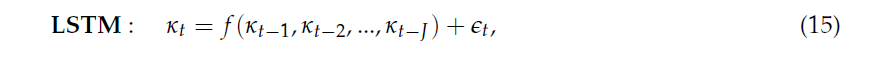
根据上述框架，我们将证明LSTM体系结构是一种出色的预测时间序列方面的工具，尤其是在时间间隔很长的情况下捕捉和管理噪音。但是像一般的NN一样，必须面对有关超参数选择的经典问题。

4.数值应用

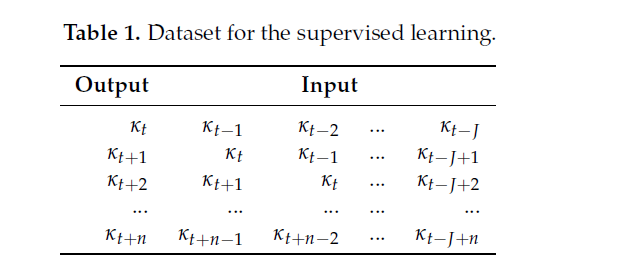
在本节中，我们将在LC模型的经典方案中介绍具有LSTM架构的RNN。 更准确地说，我们的目标是利用LSTM体系结构的优点和功能，以提高LC模型的预测能力。 为此，我们设计了一组实验，旨在测试LSTM预测未来死亡率随时间变化的性能，并将其性能与ARIMA的性能进行比较。

因此，将ARIMA模型作为预测基准时，分析将涉及时间指数kt趋势的预测，而参数ax和bx是根据Lee和Carter（1992）所应用的估计程序获得的。

与使用带有漂移随机游走的LC模型不同，我们根据Hyndman–Khandakar算法（Hyndman，Khandakar（2008））校准最佳ARIMA（p，d，q）。 在第一轮中，该过程使用适当的单位根检验检查时间序列的平稳性，并选择微分阶数d。 在第二阶段，它使用特定的信息标准（AIC或BIC）分别选择p和q的自回归和移动平均阶的最佳值。 该算法由R包预测中可用的auto.arima函数实现（Hyndman，Khandakar（2008），Hyndman等人（2019））。

然后将ARIMA性能与LSTM性能进行比较。 LSTM似乎是ARIMA（p，d，q）模型的更自然的竞争对手，因为它能够捕获序列数据内部的长期模式。 我们建立一个LSTM模型，该模型近似将κt与其时滞联系起来的函数f，如下所示：

其中J∈N是考虑的时间延迟数，e是误差项。

与其他机器学习技术一样，LSTM网络需要将数据集分为训练集和测试集。 训练集代表监督学习，而测试则用于验证模型。 监督学习数据集（表1）的构建如下：

在训练阶段之后，网络已经学会了输入-输出功能关系，并且它应该能够仅使用输入来预测κt的未来值。 实际上，输入是κt时滞的（n×J）矩阵，输出是其当前值的（n×1）向量，其中n∈N是实例数。

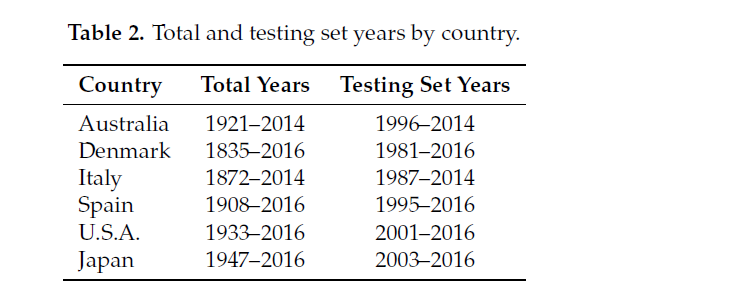
递归执行在时间n + 1，n + 2，...，n + J的κt预测值。 一般而言，使用时间为t =（n +τt 1，n +τt 2，...，n +τt J）的κt值来计算通用时间n +τ中κˆt的预测值。 输入。 重要的是要注意，在预测中，仅使用预测值κˆt而不是观测值来计算κt时间指数的未来值。 但是，尚未观察到预测数据，因此无法测量模型的准确性。

T年（对应于训练集中的最后一个观测值）是预测的起点。 因此，我们考虑T <n，其中T之前的观测值代表训练集，其余n n T形成测试集。

所有实验均使用keras，tensorflow，mrr（Bischl等人（2016）在R版本3.5.2（R统计学统计基金会，维也纳，奥地利维也纳，http://www.R-project.org/）上进行。 ），人口统计学（Hyndman（2019））和HMDHFDplus所选模型的软件包。 此外，软件包ggplot2已用于结果的可视化。

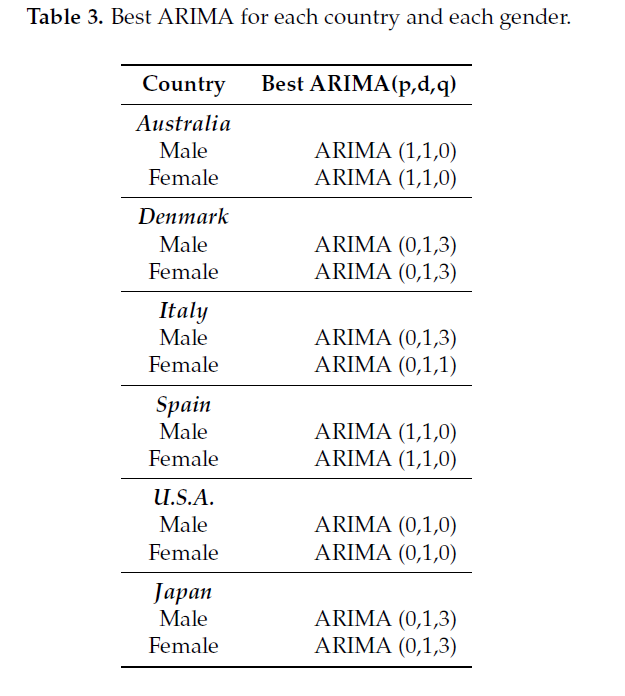
该分析是在六个国家（澳大利亚，丹麦，意大利，西班牙，美国和日本）进行的，数据取自人类死亡率数据库（HMD）。 该数据库提供了按性别区分的不同年龄和不同国家的死亡率，死亡人数和接触量。 每个国家都考虑了HMD的整个时期，并且在进行分析时考虑了一个时滞（J = 1）。

第一步是使用SVD估算LC模型参数ax，bx和κt。 提取的κt时间序列代表了分析的初始基础。 根据80％训练和20％测试的规则，将这些数据分为训练和测试集。 因此，相应地确定了观测的最后一年T。 表2中按国家列出了分析中考虑的总可用年限和相应的测试集年限。 年龄范围设置为0-100。



为了为网络选择最佳的超参数组合，针对所有国家/地区进行了初步的微调，并按性别进行了区分。 在此步骤中获得的最佳组合将用于预测过程中的LSTM校准。

关于调整的结果，我们已经注意到，在数据上，具有单个隐藏层的体系结构比其他体系结构更好，并且神经元的数量取决于国家/地区。 整流线性单元（ReLU）激活功能优于在所有国家/地区测试的其他功能。 另外，没有其他任何超参数对性能的影响的明确证据。

关于ARIMA（p，d，q）模型的校准，表3显示了按性别区分的每个国家的最佳规格。

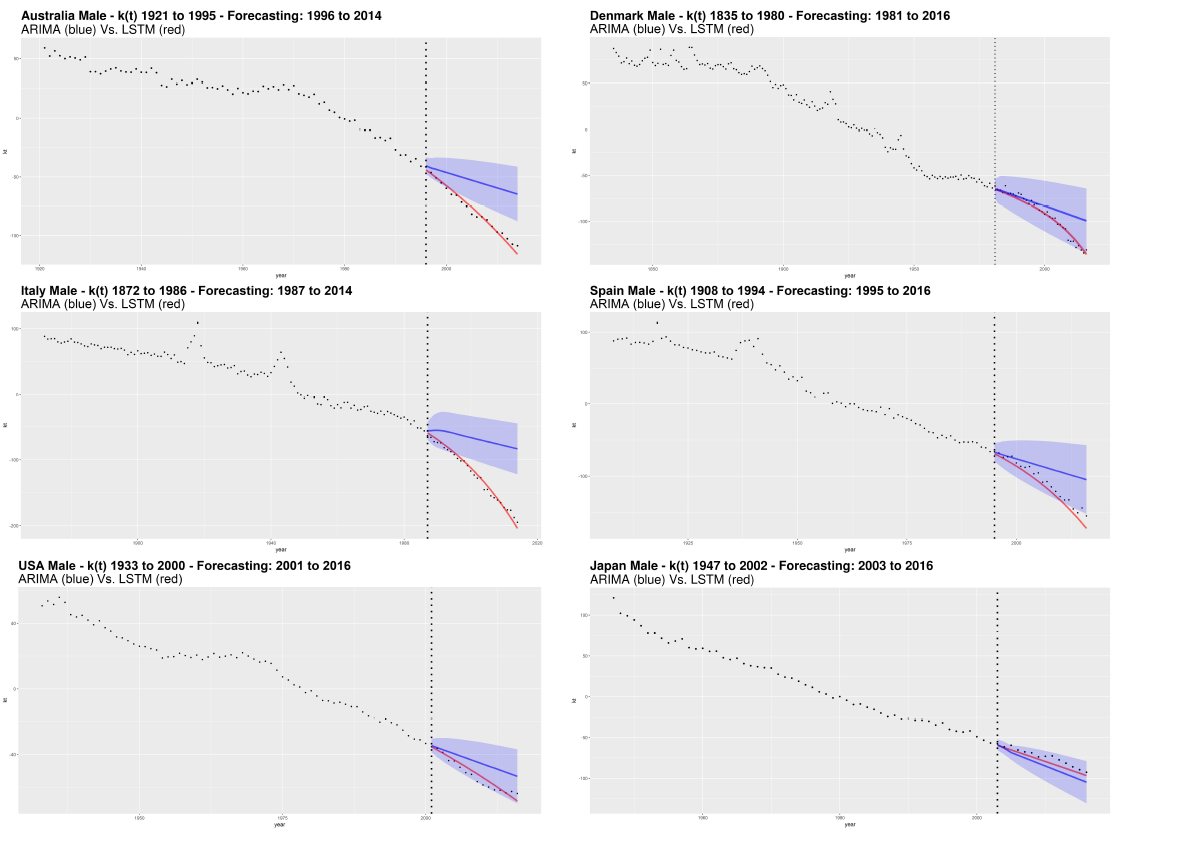
在校准步骤之后，分析包括拟合优度的数字和图形处理。 尤其是，我们遵循样本外方法，该方法代表了机器学习领域中的测试步骤。 关于图形分析，图5和图6分别显示了从SVD获得的男性和女性的κt参数估计。 垂直虚线将预测周期与用于训练LSTM网络的周期分开。 对于ARIMA模型，我们还显示了置信区间为0.995的置信度。

图5.男性人群κt的历史和预测值。

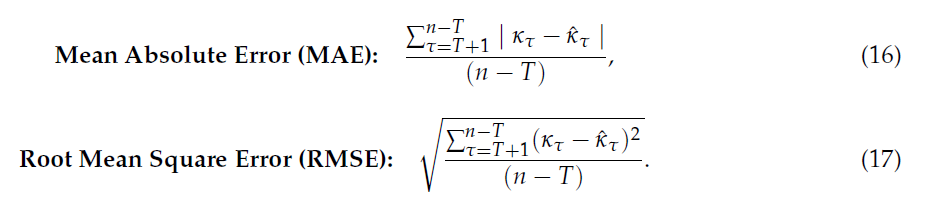
除了图形检查之外，为了比较LSTM与测试集中最佳ARIMA的性能，并测量预测质量，我们计算出以下拟合度的优度：

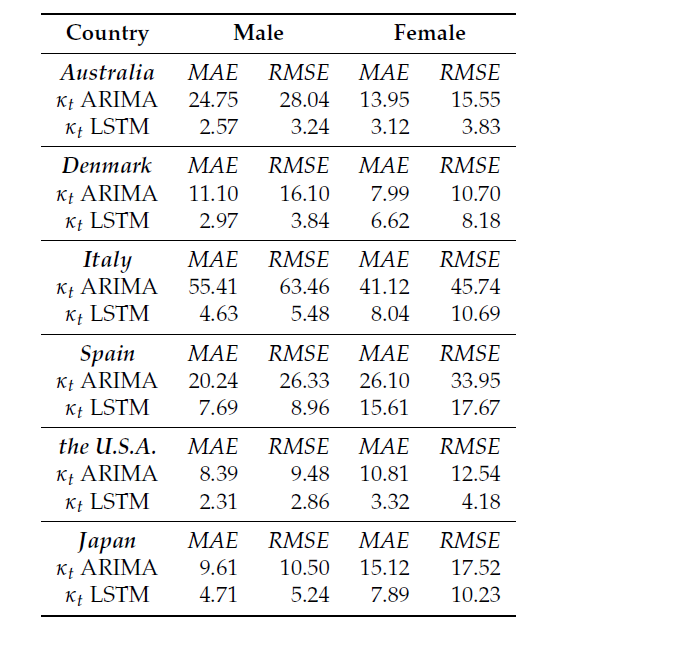
表4显示了按国家和性别划分的LSTM和ARIMA在RMSE和MAE方面的表现。

从κt图和拟合优度的结果可以看出，与ARIMA模型相比，LSTM网络提供了最佳性能。

考虑到错误度量，意大利MAE和RMSE在ARIMA方面均显示出最佳的LSTM性能，无论男女。 同样通过图形分析，LSTM似乎很好地捕捉了未来死亡率趋势的非线性，表明相对于ARIMA模型，LSTM能够更好地代表死亡率下降的动态。 日本男性和丹麦女性获得“最差”的LSTM性能，但与ARIMA相比仍然胜出。

我们注意到LSTM具有更高的捕获非线性趋势的能力，而不会在相反的情况下发生，即过度振荡或过度抛物线趋势（在传统NN中观察到后者）。 另一方面，我们的分析表明ARIMA无效。 有时根据ARIMA模型得出的κt演化甚至超出置信区间，例如意大利（男女），澳大利亚（男性）和西班牙（女性）。

表4.在每个国家的测试集中，长期短期记忆（LSTM）和ARIMA的性能。



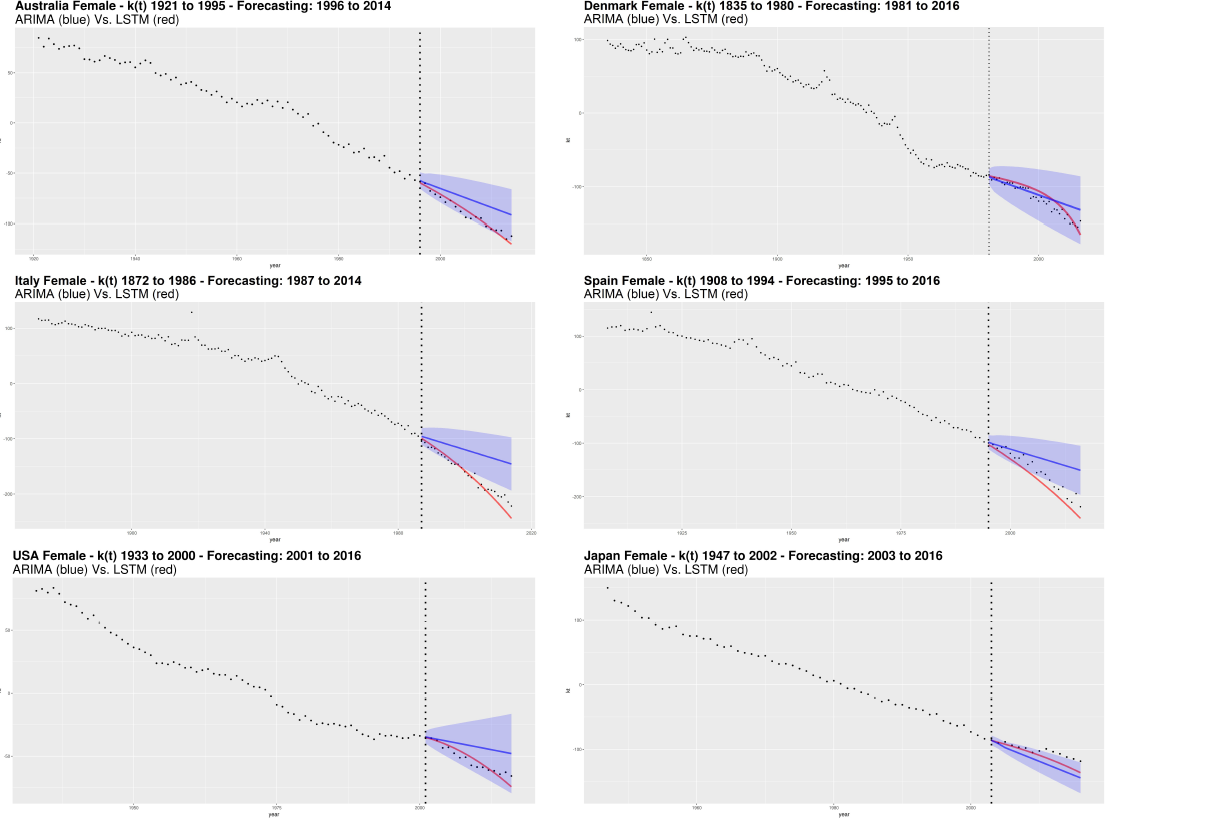


图6.女性人口的kt的历史和预测值。

我们的结果突出了ARIMA不足以检测死亡率随时间下降的动态。 尽管ARIMA过程已广泛用于对死亡率的时间指标进行建模，但它具有固定的结构，并且在数据满足ARIMA假设（例如，作为集成模型最重要特征之一的恒定方差假设）时效果很好。 在许多情况下，人口统计数据可能会显示出波动性，并且此功能不符合ARIMA的假设，尤其是对于长时间序列而言。

尽管NN是一种非常强大的学习算法，但它仅提供点预测，而没有任何迹象表明它们的可变性。置信区间预测是NN领域的真正挑战。关于这个话题的文献并不广泛，这个问题仍然是未解决的。一些学者（Khosravi等人（2011），Keren等人（2019））已经遇到了这个问题，但是到目前为止，他们中没有一个人处理有关应用于时间序列的NN（更广泛地说是深度学习）的问题。但是，LSTM网络显示出它是满足更准确地预测死亡率随时间变化趋势的理想人选。我们可以说，LSTM的体系结构可以很好地了解过去死亡率的重大影响，并可以在未来几年内以高精度替代它，因此它在所有分析的国家中都优于ARIMA模型。 LSTM网络的这种能力非常明显，尤其是对于κt参数没有明显线性趋势的人群。

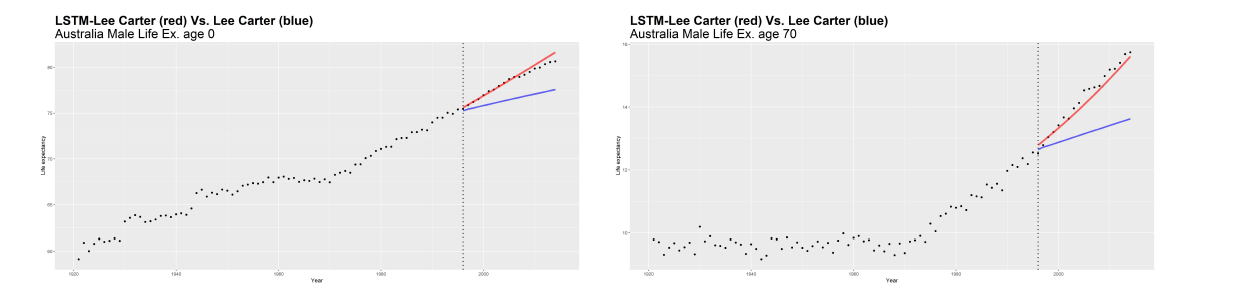
LSTM的另一个显着方面涉及在不依靠先验选择时间步长的情况下实现良好的预测性能的可能性。通过示例的方式，我们显示了通过两种模型获得的澳大利亚男性的预期寿命（图7），死亡分布Dx，t（图8）和对数死亡率log（mx，t）（图9）的预测值。此处提供的所有分析均支持预期寿命随时间明显增加的基本发现，如出生时预期寿命的情况所示（图7，左侧）。 ARIMA模型提供了简单的预测趋势形状。未来kt值的直线（随时间推移而恒定）会产生预测死亡率形状的不切实际的行为。相反，与LSTM集成的LC模型在预期寿命的预期值和实际值之间产生的差距可忽略不计。从图8和图9分别使用平滑形状描绘的Dx，t和log（mx，t）的结果中，我们可以证实，相对于ARIMA模型，LSTM对真实数据提供了更好的预测。

图7.出生时（左侧）和70岁（右侧）的预期寿命：ARIMA（1,1,0）与LSTM。

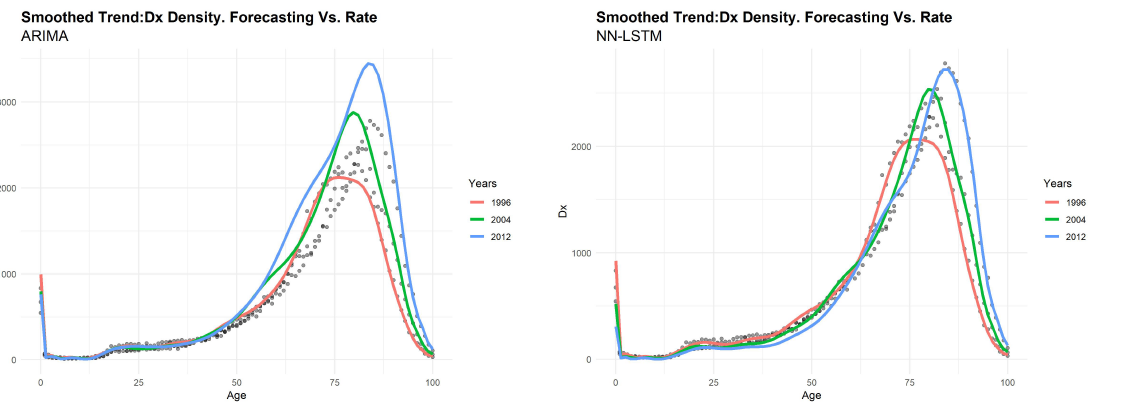


图8.死亡分布：ARIMA（1,1,0）（在左侧）与LSTM（在右侧）。 澳洲人男性人口。

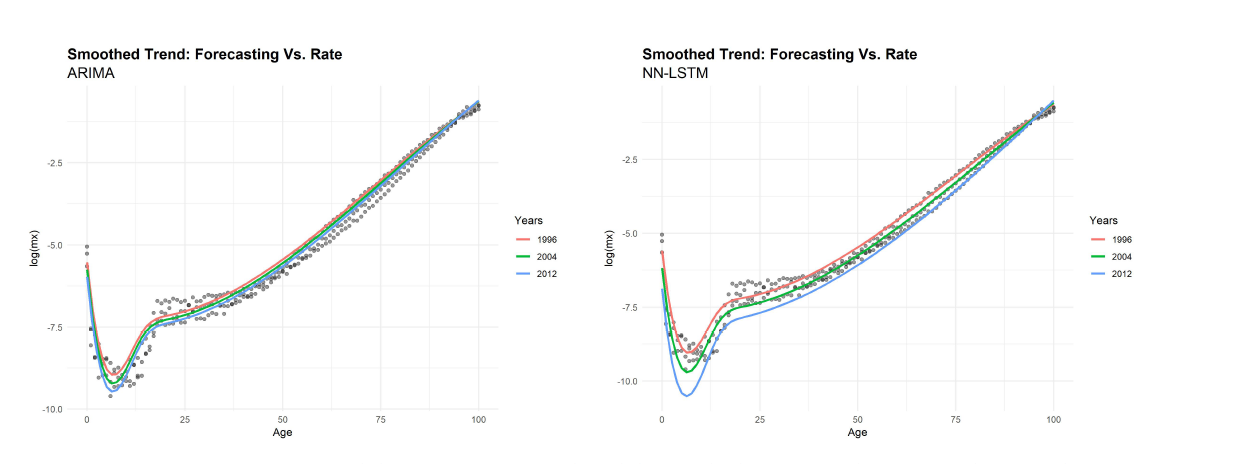


图9.中心死亡率的对数：ARIMA（1,1,0）（左侧）与LSTM（右侧）。 澳洲人 男性人口

5。结论

在死亡率领域，已经广泛提出了涉及预期寿命稳定增长的理论。 有鉴于此，我们可以断言死亡率提高是事实。 其步长仍然是一个辩论的问题。 后一点在预测方法中也有重要的体现。 因此，LC模型中随时间变化的参数的非线性估计领域应该成为一个关键点。 不幸的是，学者们一直致力于提高LC拟合度，而忽略了一个非常重要的观点：预测趋势。

在目前的工作中，我们提出了一种基于RNN和LSTM架构的深度学习集成LC模型，用于预测kt指数的未来值。 该调查已按性别在全球六个国家进行。 与规范的ARIMA相比，该方法显示了死亡率趋势预测的非常高的准确度。 实际上，由于LSTM的体系结构包括三个步骤：捕获，记忆和复制，因此LSTM可以存储出色的性能。 这些特征使LSTM可以根据死亡率随时间下降的趋势提供比最佳ARIMA方法更准确的预测。

此外，可以注意到，LSTM的能力在预期寿命趋势和死亡分布中也很明显，其中，由于采用了“样本外”方案，您可以看到一种新的思维方式对人口预测的切实影响使用创新的预测工具。经典的LC模型以及其他使用ARIMA过程的死亡率模型，都可以产生琐碎的预测趋势形状，从而随时间提供恒定的预测死亡率形状行为。 LSTM具有更逼真的非线性行为，因此可以强调随时间变化的实际值与预测值之间的差异。特别是，关于通过LSTM集成的LC模型提供的预期寿命形状，时间序列的实际值与预测值之间的差异很小。根据广泛接受的关于预期寿命不断增加的假设，我们的模型提供了一种更为乐观的死亡率假设。

该证据支持了本工作中提出的LC模型的深度学习集成的有用性，该模型提供了更为准确的预测，但同时保留了LC的简约性和鲁棒性。 生命保险行业，退休金计划和社会保障计划的从业者特别喜欢这最后一个功能，这些人必须计算和管理与寿命动态密切相关的未来现金流量。 由于LSTM（通常是NN）是一种灵活的工具，因此有可能克服ARIMA的限制并进行更实际的评估。 重要的是要记住，本文提出的方法提供了点预测，而变异性分析和置信区间的构建仍然是NN领域的实际挑战。

在未来的工作中，我们计划将分析扩展到HMD数据库中包括的其他国家，并探索LSTM在其他随机死亡率模型中的潜力。 我们认为，对LSTM（以及一般而言的机器学习模型）的可变性和鲁棒性进行分析仍然是数据科学和机器学习技术领域中最重要的要点之一（Richman和Wüthrich（2018）），这值得研究。