

基于时序自注意力机制的疫苗工艺参数预测技术

刘潇, 赵勇, 徐海涛

金宇保灵生物药品有限公司, 内蒙古 呼和浩特市 010020

摘 要 : 口蹄疫是一种极为严重的接触性传染病, 对家养、野生蹄类动物具有致命威胁, 注射疫苗能有效减少动物口蹄疫发病率。在疫苗的制备过程中, 为了精确控制生产工艺参数, 本文提出了一种基于时序自注意力机制的深度学习模型, 用于预测疫苗中核心成分 146s 抗原在最终试剂中的含量。该模型结合自注意力机制、门控残差网络和静态协变量编码器, 能够有效处理长序列数据, 并考虑外部协变量对参数的影响。本文在金宇保灵生物药品有限公司提供的数据集上进行实验, 实验结果表明该方法能较好的预测 146s 的含量, 并且相比传统 LSTM 模型具有更高的预测精度和可解释性。本研究为疫苗生产工艺参数预测提供了新的思路。

关 键 词 : 口蹄疫疫苗; 工艺参数预测; 时序自注意力机制

Vaccine Process Parameter Prediction Technique Based on Temporal Self-Attention Mechanism

Liu Xiao, Zhao Yong, Xu Haitao

The Spirit Jinyu Biological Pharmaceutical Co., Ltd, Hohhot, Inner Mongolia 010020

Abstract : Foot-and-mouth disease (FMD) is an extremely serious contagious disease, posing a lethal threat to domestic and wild cloven-hoofed animals. Vaccination can effectively reduce the incidence of FMD in animals. In the process of vaccine preparation, to accurately control the production process parameters, this paper proposes a deep learning model based on temporal self-attention mechanism for predicting the content of the core component 146s antigen in the final reagent of the vaccine. The model combines self-attention mechanism, gated residual networks, and static covariate encoders, which can effectively handle long sequence data and consider the impact of external covariates on parameters. Experiments were conducted on the dataset provided by the Spirit Jinyu Biological Pharmaceutical Co., Ltd, and the results show that this method can predict the content of 146s relatively well, and compared with the traditional LSTM model, it has higher prediction accuracy and interpretability. This study provides new insights into the prediction of vaccine production process parameters.

Key words : Foot-and-mouth disease vaccine; process parameter prediction; temporal self-attention mechanism

引言

口蹄疫是一种对牛、羊、猪等蹄类动物极具破坏力的传染病, 其爆发常导致畜牧业和相关经济遭受严重损失。根据现有口蹄疫研究来看, 注射灭活疫苗是预防和控制该疾病的主要手段之一^[1]。随着生物技术的不断发展, 基于细胞培养的口蹄疫疫苗生产技术逐渐成为主流。由于病毒具有极强的宿主性, 病毒的成长完全依赖于原宿主细胞的能量和代谢系统, 所以宿主细胞的好坏直接影响到产品的质量。但是现有的疫苗生产工艺通常基于经验和试错来确定关键参数, 这种方法存在着效率低下、周期长、成本高等问题。为减小人工经验对疫苗制备工艺的影响, 基于时序自注意力机制的人工智能预测技术逐渐成为了疫苗工艺参数预测的研究热点。时序自注意力机制是一种基于神经网络的新兴技术, 其结合了自注意力机制和时间序列建模的方法, 能够有效地捕捉数据中的时序信息和内在关联, 从而实现对复杂系统的精准建模和预测。

* 作者简介:

刘潇, 男, 蒙古族, 硕士学历, 高级工程师, 研究方向为微生物发酵、制药工程

赵勇, 男, 汉族, 硕士学历, 中级工程师, 控制系统与智能控制、自动化设备与系统应用

徐海涛, 男, 汉族, 本科学历, 研究方向为信息化设备与系统

致谢: 该论文成果受内蒙古自治区科技计划项目(编号: 2022YFXM0005)资助

一、研究背景与基础

（一）疫苗制备工艺

转瓶培养是目前动物疫苗生产的一种关键技术，然而，这种培养方式也存在显著缺陷，如不同转瓶间的细胞质量和病毒产量存在显著差异，这直接影响到疫苗批次间的均一性和效价的稳定性。此外，该过程对工人的劳动强度要求较高，同时，转瓶内潜在的污染源可能导致无法预测的高内毒素问题，国内许多生产企业也在生产技术上创新，推出了悬浮培养这一新型培养模式。悬浮培养不依赖于细胞贴壁生长，而是让细胞悬浮在液体培养基中进行生长。相较而言，由于微载体系统的均匀性，细胞在培养过程中能够充分接触到养分和氧气，能够实现更高的细胞密度和病毒效价；微载体体系中细胞均匀地附着在微载体表面，从而实现更高的细胞密度更好的疫苗质量；自动化的细胞培养和收集过程可以降低能耗和劳动力需求^[2]。

梅建国等^[3]利用 Cephodex 微载体进行伪狂犬病疫苗的生产，与传统的转瓶培养方式相比，Cephodex 微载体能够提高细胞的生长速度和病毒的产量，从而增加了疫苗的效价和稳定性。同时，Cephodex 微载体还能够降低工人的劳动强度，减少了生产过程中的污染风险。解占长^[4]对疫苗的生产工艺进行了优化，通过大规模培养和纯化，获得了高纯度的重组腺病毒疫苗。同时，为了评估疫苗的安全性和有效性，他还建立了一种动物模型替代方法，利用了小鼠等实验动物，通过模拟 PRRSV 感染过程，表明欧洲型 PRRSV 重组腺病毒疫苗具有良好的安全性和免疫原性。龚迪等^[5]通过研究 MDCK 细胞在微载体悬浮培养中的放大工艺，钙和镁离子对于细胞贴壁是不可或缺的，而麦麸水解物可以部分代替培养基中的血清，为工业规模反应器中微载体悬浮培养 MDCK 细胞生产流感病毒奠定了基础。

综上所述，微载体悬浮培养技术是一种更先进更有效的疫苗生产方式，更适合于本文中口蹄疫疫苗病毒的培养。

（二）疫苗生产工艺参数

疫苗的生产过程涉及到多个复杂的生物化学反应，这些反应的进行和调控都离不开各种工艺参数的控制。疫苗制备过程中常见的工艺参数包括细胞密度、PH值、温度、流量等结构数据、以及部分非结构数据。这些参数的精确控制对于疫苗的生产至关重要，工艺参数的波动和不确定性会对疫苗的质量和产量产生影响。例如，过高的细胞密度会导致细胞之间的竞争加剧，影响细胞的生长和代谢；pH值的变化可能会影响疫苗的稳定性和安全性；温度的波动可能导致疫苗中的活性成分失活或降解，从而影响疫苗的效果等。因此，对工艺参数进行精确预测和控制，是疫苗生产过程中的关键问题。

（三）时序预测

时序预测是机器学习中一种重要的数据分析任务，通常用于分析过去的模式或趋势，预测未来一系列时间步骤的数值或者事件。时序预测在当今社会扮演着重要的角色，被广泛应用于金融、气象、交通和能源等领域。其为决策者提供了有价值的参考信息，帮助人们做出更明智的决策。随着神经网络的发展和计

算能力的提升，深度学习模型如长短期记忆网络（LSTM）、门控循环单元（GRU）和卷积神经网络（CNN）等在时序预测中展现出了很强的能力。这些深度学习模型能够更好地捕获时间序列数据中的复杂关系和模式，从而提高预测的准确性。LSTM 和 GRU 等适用于长期依赖关系的模型，能够有效地处理时间序列数据中的记忆和信息传递；而 CNN 则在一维卷积的基础上，能够有效地提取时间序列数据中的局部模式。除了传统的深度学习模型，一些新兴的架构和技术也在时序预测中得到了广泛应用，如自注意力机制和生成对抗网络（GAN）等。这些模型和技术的不断发展使得时序预测在各个领域中都取得了显著的进展，并为未来的预测工作提供了更多可能性。然而，目前的深度学习方法在时序预测中仍存在三个普遍问题：第一，深度学习模型通常需要大量的数据进行训练，特别是在处理复杂的时序数据时。如果数据量不足，模型可能会过拟合或者无法捕捉到数据规律。第二，深度学习模型通常被认为是黑盒模型，难以解释其中的预测过程和决策根据。在某些场景下，特别是需要理解预测背后原因的情况下，缺乏解释性可能会导致时序预测做无用功。第三，尽管 LSTM 等模型能够一定程度上处理长期依赖关系，但在某些复杂的时序数据中，特别是疫苗生产工艺的复杂数据中，过多的影响因素和时间序列会严重干扰模型的长期依赖捕捉能力，导致预测精度下降。为了解决上述问题，本研究以自注意力机制为基础，构建了一种可解释性的时序自注意力机制网络，详细介绍了网络中各模块的构成和作用，以及实验中的参数设置，并结合疫苗制药公司提供的数据进行了实验。实验结果表明，与传统的 LSTM 模型相比，本研究提出的时序自注意力机制网络在预测疫苗工艺参数时更加精确有效。

二、实验数据与实验环境

（一）实验数据

本次实验数据为自建数据集，数据来源为金宇保灵生物药品有限公司 2022 年 6 月至 2023 年 12 月期间生产口蹄疫疫苗抗原 146S 过程中，所测得的抗原液各项数据。该公司通过运用智能传感技术于生物过程领域，高效地收集生产过程中的关键数据，包括转出细胞密度、纯化剂添加及消耗量、抗原 PH 值、浓缩液浊度等数据。这些数据共同构成了一个详尽的细胞生产过程大数据集，并最终被整合到一个基于 SQL 或 Oracle 的数据库系统中。

（二）实验软硬件环境

为了实现 TFT 模型的训练并完成对疫苗工艺参数的预测，需要完成实验所需的软硬件平台的搭建。在实验所需的硬件方面，本次使用了一台搭载 Linux 操作系统的远程服务器，并利用其搭载的 4 块 RTX 2080ti 图形处理器（GPU）进行模型，其强大的计算能力为加速模型训练和推理速度提供算力支持。实验的软件配置中，选择 TensorFlow 框架并在其基础上进行环境搭建。实验需要在 Linux 系统中安装 CUDA 和 cuDNN，并通过 Python 语言进行程序编写。CUDA 是 NVIDIA 推出的并行计算平台和编程模型，cuDNN 是针对深度神经网络的加速库，安装这两个库可以确

性的比率，如下所示。

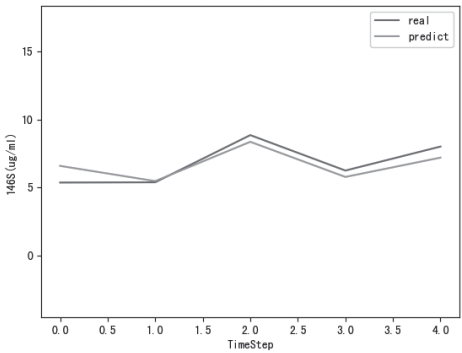
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{MSE}{Var} \tag{4}$$

其中 Var 为已知数据的方差。

表1 TFT 与 LSTM的各项回归评价指标

	MSE	RMSE	MAE	MAPE	R2
本文方法	0.505105	0.710707	0.597027	0.083888	0.747415
LSTM	5.891764	2.427296	1.319458	0.134126	0.215324

使用本文方法与 LSTM方法在实验数据集上进行实验的结果如表1所示。通过实验数据可以发现利用本文方法对口蹄疫病毒病毒培育数据集进行预测后，146s的拟合程度普遍高于普通时序预测方式 LSTM进行预测的结果，且最终决定系数达到了0.75，能够较好的解释细胞密度、细胞类型等因素对浓缩 146s的影响。如图3所示为最终可视化后的浓缩 146s拟合曲线。从图中可以看出，本文所提出的方法能够较好地拟合真实数据，证明了本文方法的有效性。



>图3 浓缩 146s拟合曲线图

五、结语

为了解决疫苗生产工艺中浓缩 146s参数预测问题，本文设计了一种时序自注意力机制网络。基于自注意力机制设计了可解释性的时序自注意力层，提高了网络模型对复杂时序数据中长期依赖性的捕捉能力，并能输出各个疫苗工艺参数对预测目标的影响。该网络还结合了门控残差单元用于筛选出重要输入数据，以及静态协变量模块用于嵌入时不变信息，可以最大化利用所有输入的时序数据。通过在疫苗制药公司合作提供的数据上进行的实验可以表明，本文设计的网络在复杂时序数据上的预测能力大于传统的时序预测方法，且具有可解释性的输出能够为决策提供更多的参考信息。

参考文献

[1] 谭涛,任静雷,倪春霞,等.不同抗体检测方法对猪口蹄疫疫苗临床免疫效果的评估[J].中国猪业,2022,17(03):66-69.
[2] 邵三敏,吕宁宁,马冬,等.动物疫苗生产工艺研究的现状分析[J].中国动物保健,2024,26(03):104-105.
[3] 梅建国,庄金秋,管宇,等.应用 Cephodex 微载体规模化生产伪狂犬病疫苗的研究[J].中国预防兽医学报,2016,38(03):240-244.
[4] 解长占.欧洲型 PRRSV 重组腺病毒疫苗生产工艺及动物模型替代方法建立[D].吉林农业大学,2017.
[5] 龚迪,易小萍,张元兴.MDCK 细胞微载体悬浮培养放大工艺研究[J].中国生物工程杂志,2012,32(09):55-60.
[6] Lim B, Arik S Ö, Loeff N, et al.Temporal fusion transformers for interpretable multi-horizon time series forecasting[J].International Journal of Forecasting, 2021, 37(4): 1748-1764.