

基于 ADS-B 数据的航迹预测初探

陈汀

民航华北空管局, 北京 100621

摘要: 近年来, 人工智能和大数据领域的发展取得了显著的成就, 尤其是在卷积神经网络技术方面的应用效果更是不容小觑。ADS-B 轨迹数据具有独特的多维时空特性和显著的不确定性。本文的研究主要包括: 在时序特征的建模上, 针对于轨迹预测的研究中, 探讨了采用不同的长短期记忆网络 (LSTM) 的航迹预测方法。

关键词: 广播式自动相关监视; 轨迹预测; 长短期记忆网络;

作者简介: 陈汀 (1983-), 男, 汉族, 江苏, 硕士研究生, 助工, 研究方向: 人工智能大数据

近期, 国内航空业进入了迅猛发展的阶段, 便利且迅速的旅行方式已经成为众多乘客出行的首选。展望未来, 民用航班和无人机的数量都将实现迅猛增长, 预计在未来十年内, 航空器的总数将有可能达到目前的两倍, 这无疑会给空中交通流量的管理带来更大的压力。正在进行实验的下一代空中交通管制系统 TBO (基于航迹运行), 该系统的设计理念是基于初始计划的信息, 以 4D 航迹管理为核心, 专注于对航空器的 4D 航迹进行精确计算和预测。这一理念已经被国际民航组织 (ICAO) 纳入全球航空导航计划 (GANP), 被认为是提升空中交通管理效率的关键要素^[1]。针对 4D 航迹的预测研究的热度逐渐上升, 国内外的研究也很多。文献^[2]提出了一种基于 BADA 数据库的 4DT 航迹预测模型, 该模型从航班的高度, 速度, 航向角等方面重构了约束条件, 提出 RTA 优化了 4DT 算法。

2019 年马兰, 高永胜在文献^[3]中提出了基于 ADS-B 数据挖掘的航迹预测方法, 该方法对原始数据进行降噪等处理后使用了 CURE 聚类分析。董欣放在文献^[4]中提出了基于高斯混合模型的终端区 4D 航迹建模和预测。Daeke 在文献^[5]中提出并证明了 S 模式雷达和 ADS-B 数据的航迹数据能够动态预测 (DAP) 航空器下传参数。Yu Liu 在文献^[6]中发现通过 ADS-B 的实时航迹数据, 可以更新飞行航迹的航向角方向, 提出了 IMM 交互式算法预测飞行器航迹。针对时间序列数据的预测, 深度学习尤其集中于长短期记忆

网络 (LSTM), 是因为该网络在相关任务中表现出色, 因此目前的研究主要聚焦于 LSTM 网络的应用。

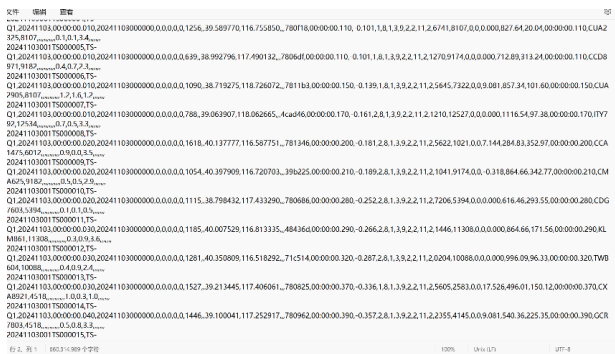
随着诸如 ADS-B (广播式自动相关监视) 和 TBO (基于航迹运行) 等新兴技术的应用, 空中交通管理系统产生了海量数据, 提炼出这些信息的关键价值成为民航实现高质量发展的重要基础与挑战。当前, 在计算机视觉处理、自然语言处理以及语音识别等多个领域, 深度学习均展现出了良好的性能。不少科研人员采用深度学习技术分析 ADS-B 数据集并建模, 以实现轨迹预测, 这一方向近年来逐渐受到关注。本文借助于 ADS-B 数据中心的 ADS-B 数据集, 进行了一系列针对航迹预测的深入探讨与分析, 旨在为航迹流量的预测提供新的视角与思路。

一. ADS-B 数据集的构建

1.1 ADS-B 数据的收集

ADS-B 数据中心系统是一种全天候运行的系统, 具备高稳定性和可靠性, 旨在提供实时的 ADS-B 数据处理, 以保障空中交通的安全。该系统的架构采取了分布式设计, 基于客户机 / 服务器模式进行实现。目前 ADS-B 数据中心所能接收到的数据主要包括有关航空器的信息, 其内容主要涉及航班编号、24 位地址码、经度和纬度、高度、速度以及航向等特征。这些数据特点均可用于后续的数据预测模型计算。在这一过程中, ADS-B 解码装置负责对接收到的消息进行处理, 成功解析的消息将以 CAT021 的格式被保留在本地数据中心。该中心会对原始信息进行多次整理, 并将提

取出的数据整理成 .txt 格式，以对应的日期命名并保存于服务器，为后续分析与挖掘真实的 ADS-B 数据做好准备。这些信息被存储在一个称为 FDP（飞行数据处理器）的专用服务器上，该服务器每小时会记录一次日志文件。该日志文件中的数据主要来源于各地的地面站发送的航空器信息，而这些 ADS-B 相关数据无法直接被使用，因为发送的内容格式为 .txt 文件。



1.2 数据预处理

数据清洗作为数据预处理的核心环节，主要用于解决数据采集过程中因系统故障、人为失误或传输错误等因素引发的数据质量缺陷。其核心目标在于通过系统化技术手段，在保持以采集的数据集的真实性以及其结构完整的基础上，对缺失记录和冗余记录以及无用数据进行处理，从而消除数据噪声对分析过程的干扰。为了获得特定航班的全面飞行轨迹数据，初步步骤是运用 pandas 库将 .txt 文件转换为 CSV 格式，并从中提取各个日期的航班信息，接着以航班号为基础进行分组，同时生成以航班号命名的 CSV 文件。

研究人员对多条航班路径进行了比较，最终选定了一条数据记录最为详尽的航班轨迹。该航班的数据包含了完整的航行路线和所有重要的航路点，并涵盖了起降的机场。这一选择将为随后的神经网络训练提供充分的支持，并最大程度地降低因数据不足造成的性能评估误差。经过分析文件中的关键数据，我们得到了某航班的飞行路径信息。接下来，还要继续处理这个数据集。首先，删去冗余字段。接着，将日期与时间这两项索引整合为一个统一的时间索引，这种方法

可以减少冗余的时间数据，便于后续的分析 and 处理。若数据集中存在少量的缺失值，可以采用前一时刻和后一时刻的数值进行平均，来进行替换，以此来降低测算误差。接着，利用线性函数的归一化方法对每一列的数值进行处理，以确保数据在同一量纲上。最终，航班的轨迹数据被分为训练集和测试集。选择航班数据的前 75% 的部分作为模型的训练数据；将剩余 25% 的数据当作测试集，旨在为了进行预测，并借此计算不同算法之间的误差。将时序数据转换成监督学习形式，通过定义历史视窗的时长、预测的时间步及输出的维度，可以有效地对时序特征进行工程化处理。以一个原始的 $N \times 6$ 维数据集为例，其中包含时间戳及五个特征如经度、纬度、高度、航向角和垂直速度，若采取前两时间步的数据来预测接下来的一个时间步，将会形成一种样本构造方式。这种方法会使用连续的两帧 6 维数据作为输入，并将下一帧 6 维数据作为输出，从而通过滚动窗口的设计来确立时间序列之间的关联性与特征间的耦合关系。在一个包含了多个时间序列的数据集里，我们通过尝试设定每个时间步只有两个观测值，例如仅考虑经度和纬度，如下图所示；

Time	Latitude	Longitude
58:56.1	39.54052	116.3919
58:57.1	39.54052	116.3919
59:00.0	39.54125	116.3918
59:01.0	39.54125	116.3918
59:29.0	39.54125	116.3918
59:30.1	39.54166	116.3917
59:33.0	39.54166	116.3917
59:34.0	39.54263	116.3915
59:39.0	39.54263	116.3915
59:40.0	39.54263	116.3915
59:43.0	39.54308	116.3914
59:44.0	39.54308	116.3914

表 1-1 经纬度数据集

二、基于不同的 LSTM 的轨迹预测方法

本节的实验对三种时间序列预测模型进行了比较

分析, 这些模型分别包括基础的单层长短期记忆网络 (LSTM)、双向长短期记忆网络 (Bi-LSTM) 以及多个隐层堆叠的长短期记忆网络 (Stacked LSTM)。通过评估不同模型架构的性能, 分析其预测能力的差异。

经典的 LSTM 通过引入记忆单元和门控机制, 成功缓解了循环神经网络在处理长距离依赖关系时所面临的挑战。然而, 它存在局限, 仅具备单向时序处理的特点, 因此在推导当前状态时, 仅依靠过去的信息, 而不考虑未来信息的影响。因此, 双向长短期记忆网络 (Bi-LSTM) 应运而生, 通过双向的时间序列建模来融合以往与未来的信息关系, 从而增强模型的上下文理解能力。Bi-LSTM 的核心由两个对称且参数互不干扰的 LSTM 网络构成, 其中正向 LSTM 按照时间顺序处理输入序列, 积累从历史到当前的依赖特征, 而反向 LSTM 则以逆序的方式扫描序列, 以捕捉从未来到当前的潜在联系。Bi-LSTM 在参数的调整上与传统 LSTM 相似, 在进行反向传播时, 需要对双向网络的权重进行同步优化。

2.1 基于 LSTM、Bi-LSTM、Stacked LSTM 的预测模型

该单层 LSTM 模型由四个主要部分组成, 其中输入层的时间步长被设置为 16。在此配置下, 每个样本由 16 个连续的时间步构成, 每个时间步包含 6 个不同的特征变量, 形成 16×6 的输入数据维度。其次, 采用 LSTM 层, 该层设置 50 个隐藏单元, 旨在挖掘时间序列中的动态特征。其三为正则化层, 紧接在 LSTM 层之后, 采用 Dropout 方法, 随机舍弃 20% 的神经元, 以防止模型陷入过拟合的困境。其四为完全连接层, 此层功能为将 LSTM 输出的高维特征转化为单一预测值, 维度设定为 1, 以适应后续单变量的时间预测任务。为提升模型的泛化能力, 所有的隐藏层输出都经过了 Dropout 的处理。值得一提的是, LSTM 层在时间维度上能够产生完整的序列结果, 但最终预测时仅依赖最后一个时间步的输出, 这样可以更好地强调序列末尾特征与预测目标之间的密切关系。

Bi-directional LSTM 模型是一种单一结构的设

计, 其核心在于用 Bi-directional LSTM 取代传统的 LSTM 单元。此模型包含一个输入层、一个双向 LSTM 层、一个丢弃层以及一个全连接层。

堆叠式长短期记忆网络是由多个 LSTM 单元构成的复合模型, 我们称之为堆叠式 LSTM 模型。该模型通过交替堆叠 LSTM 单元与全连接层来实现结构的构建。该结构设计包含一个输入层、两个 LSTM 的隐含层、两个 dropout 层以及最终的全连接层。

2.2 预测任务的设计

本研究将采用 LSTM、双向 LSTM (Bi-directional LSTM) 以及堆叠式 LSTM (Stacked LSTM) 三种类型的模型, 专注于对 71bf55 航班的数据集进行轨迹预测, 并展开性能的评估。在数据处理阶段, 将该航班的总数据量的前三分之二分配为训练集, 而剩余的四分之一则用于测试集。在将时间序列转换为监督学习任务的过程中, 采用特定的转换方法。模型以前 16 个时间步的经度、纬度、高度、航向角和垂直速度五个维度的数据作为输入, 并将第 17 个时间步的经度作为输出目标。随后, 输入从第 2 步到第 17 步的六个维度数据, 而对应的输出则是第 18 步的经度, 依此类推。通过这一系列循环操作, 整个训练数据得以有效地转换为清晰的输入输出格式, 从而为模型学习潜在的数据模式提供了极大的便利。依据以上方法, 可以将轨迹预测视作对单一变量在短时间内的逐步预测, 旨在对经度进行有效的预估。关于其他变量的预测, 其基本原理与经度的预测方法相似, 因此在此不再详细阐述。

在此, 我们定义 (P_i) 为在时间 (i) 时刻所预测的轨迹, 而 (R_i) 则为该时刻的实际飞行路线。鉴于评估模型预测性能的相关指标, 如果我们关注的这三个数值 (RMSE、MAE、MAPE) 越接近于零, 那么说明所预测的值与实际轨迹数据之间的差异也就越小。这种情况充分表明所使用的模型在预测飞行轨迹时展现出了更高的准确性。

三. 实验结果与性能测试

3.1 为了清晰展示各个模型对于航班飞行路径的

预测能力, 计划对测试集中前 2500 个数据点 (约 500 秒) 进行模拟。每幅仿真图都将同时展示真实的航迹数据与三个模型提供的预测结果, 以便于直接对比各个模型的预测效果。综上所述, 通过对不同模型的均方根误差 (RMSE)、平均绝对误差 (MAE) 及平均绝对百分比误差 (MAPE) 进行汇总, 可以从误差的角度对各模型的预测效果进行深入评估, 不同特征值的预测如下图所示:

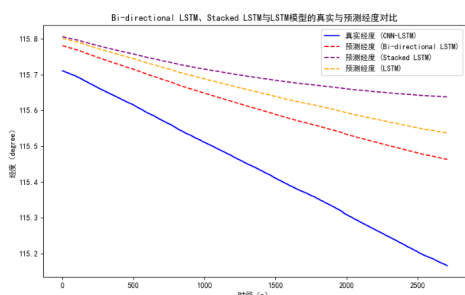


图 3-1 经度预测结果

Fig. 3-1 Longitude prediction result

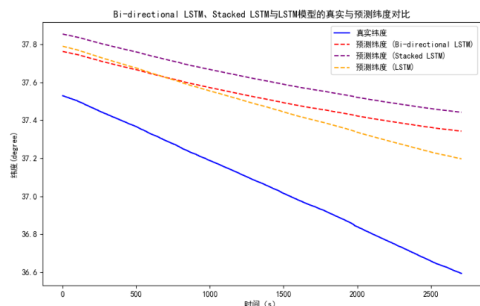


图 3-2 纬度预测结果

Fig. 3-2 Latitude prediction result

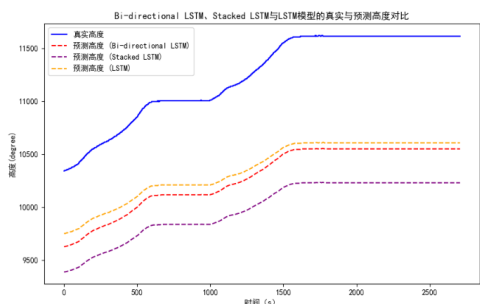


图 3-3 高度预测结果

Fig. 3-3 Altitude prediction result

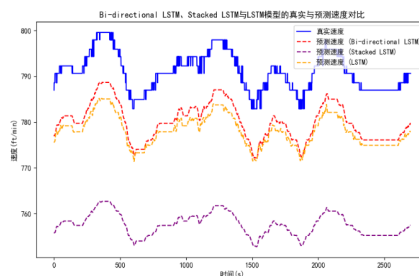


图 3-4 速度预测结果

Fig. 3-4 Speed prediction result

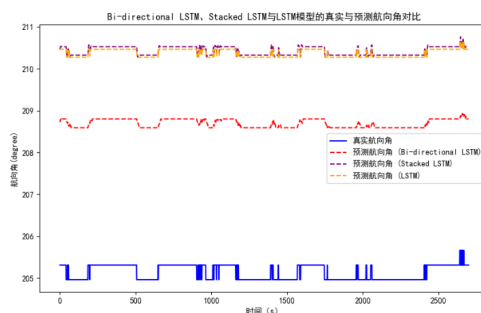


图 3-5 航向角预测结果

Fig. 3-5 Track angle prediction result

3.2 实验性能分析

通过对上述实验结果的模拟分析, 涉及的参数提示出, 当应用长短期记忆网络 (LSTM)、双向长短期记忆网络 (Bi-directional LSTM) 以及堆叠长短期记忆网络 (Stacked LSTM) 来预测五个特性变量时, 仅在经度与速度这两个特性变量的预测结果上, 其曲线与实际轨迹的形态相对接近; 而在纬度、高度以及航向角等三个特征变量的预测中, 则显现出与实际值存在显著差距。尽管上述三种模型在某种程度上能够实现预测功能, 但总体效果仍显不足。随后, 将通过计算均方根误差 (RMSE)、平均绝对误差 (MAE) 和平均绝对百分比误差 (MAPE) 这三项性能指标来进一步验证结果。RMSE 代表的是标准误差, 而 MAE 则指代

平均绝对误差，这两者均为衡量误差的方式，其数值越小表示性能的优越性越高。

根据计算结果，三种模型在经度预测方面表现最佳，而在航向角预测中则显得最为不理想。此外，这三种模型在航向角和速度的预测能力上，都显著逊色于对其他三个特征变量的预测表现。对五个特征变量预测的均方根误差（RMSE）进行求和并计算平均之后，可以得出 LSTM 的平均预测误差为 144.70，双向 LSTM 为 171.25，而堆叠 LSTM 则为 306.69。由此可见，LSTM 的预测能力明显优于双向 LSTM，且表现也优于堆叠 LSTM。如下表所示：

Data 特征	Bi-directional LSTM			Stacked LSTM			LSTM		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
经度	0.1253	0.1192	0.1033	0.2251	0.2097	0.1818	0.2508	0.2301	0.1995
纬度	0.4699	0.4539	0.8271	0.5321	0.5209	1.4076	0.5047	0.4871	0.6169
高度	837.71	834.78	7.4155	1501.41	1485.43	13.1647	698.5	693.05	6.147
速度	14.7499	14.7159	1.8601	27.0243	26.9683	3.4085	16.1974	16.1654	2.0433
航向角	3.2029	3.2014	1.5607	4.2657	4.2647	2.079	8.0556	8.055	3.9268

表 3.1 三种模型预测误差

Tab. 3.1 Three models predict errors

平均绝对百分比误差通常称为 MAPE，当计算结果为 0% 时，可以认为对应的模型是完美无缺的；相反地，如果计算结果超过 100%，则说明该模型的效果较差，可能与真实值存在较大偏差。通过相关的分析发现，对于经度和纬度的 MAPE 值计算，LSTM 和双向模型所得到的结果均低于 1。这表明，这两个模型在这三项变量上所作的预测是值得参考的。然而，在预测经度特征时，Stacked LSTM 模型的结果才显得比较有意义。

总体而言，以上提到的三个模型在某种程度上具有一定的预测能力，然而它们的表现仍有待增强。在不同特征的预测任务中，三种模型对经度的预测表现较为理想，但在其他特征的预测上，则尚有显著提升的空间。

四 结语

对比三种 LSTM 预测模型中，LSTM 模型的表现相对较好。经过具体实验的验证，简单地将 LSTM 进行堆叠而构建的 Stacked LSTM 模型并未提升性能，反倒出现了性能下降的情况。由此可知，提升预测效果

的关键在于对网络模型进行巧妙而细致的设计。在对速度进行预测时，三种模型中，双向 LSTM 展现了更为突出的性能，并且整体预测效果上，它与 LSTM 模型的差异也并不显著。综上所述，在分析 LSTM、Bi-directional LSTM 与 Stacked LSTM 三种模型的预测结果时，在不同特征值方面均表现出了各自的优势和不足的地方。

基于 ADS-B 历史数据挖掘的航迹预测方法，适用于实时性较高的空域管理系统，其预测航迹可为 4D 计算提供参考。结果表明，该方法能较为准确预测出 4D 航迹，对实际飞行具有一定参考作用。

参考文献：

[1] 李萌. TBO 模式下航空器四维航迹规划技术[D]:[硕士学位论文]. 河北：华北理工大学，2022.

[2] Zhiming Ma, Mengfei Yao, Tao Hong, Bo Li. Aircraft Surface Trajectory Prediction Method Based on LSTM with Attenuated Memory Window[J]. Journal of Physics: Conference Series, 2019, 1215(1):66.

[3] 王涛波，黄宝军. 基于改进卡尔曼滤波的四维飞行航迹预测模型[J]. 计算机应用，2014, 34(6): 1812-1815.

[4]Kastella K, Biscuso M. Tracking algorithms for air traffic control applications[J]. Air Traffic Control Quarterly, 1995, 3(1): 19-43.

[5] 马 兰，高永胜. 基于 ADS-B 数据挖掘的 4D 航迹预测方法[J]. 中国民航大学学报, 2019, 37(4):1-4.