



華東理工大學

机器学习工具与平台 课程设计

2024-2025 学年第 2 学期

专业:	人工智能
班级:	人智 222
学号:	22013349
姓名:	胡佳铭

20 25 年 6 月 20 日

高安市天气预报

作者， 胡佳铭

华东理工大学

摘要：通过获取到的五年内的天气、气压、气温的气象学基本特征数据，依照时间步的划分从而进行机器学习以及深度学习的预测，实现以天为单位气温、湿度的回归预测与天气的分类预测。

关键词：深度学习、机器学习、气象预测

引言：在当今社会，气象预测在农业生产、防灾减灾、交通运输和人们日常生活中扮演着至关重要的角色。随着人工智能和大数据技术的飞速发展，利用计算机模型对未来气象状况进行准确预测成为可能。本研究围绕高安市地区展开，基于收集到的历史气象数据，构建了一个气象预测模型，旨在对该地区的未来天气状况、气温变化以及湿度水平进行有效预测。通过模型的建立与验证，不仅有助于提升当地对气候变化的响应能力，也为相关决策提供科学依据，具有较高的实用价值和应用前景。

1 实验目的

本实验旨在构建一个能够准确预测高安市未来天气状况的气象预测模型，重点包括对天气类型、气温变化趋势及空气湿度水平的预测。通过对历史气象数据的采集与分析，结合合适的机器学习或时间序列建模方法，挖掘气象因子之间的内在关联规律，提升气象预测的精度与实用性。本实验不仅验证所选模型在特定地区的适用性与泛化能力，还为今后在更大范围内开展区域性气象预测研究提供方法参考与技术支持。

2 数据来源

本实验数据来源于rp5.ru对天气记录的整理，数据图表（图一）如下：

当地时间	T	Po	P	Pa	U	DD	Ff	#f10	#3	N	WW	W1	W2	Tn	Tx	Cl	Nh	H	Cm	Ch	VV	Td	RRR	IR
23.06.2022	23.9	749.5	753.7	-0.9	94	从东方吹来	1			4	间歇性雨，在整个相时在整体相时	23.0	25.0				2500或更高，或无云。	5.0	22.9	13	12			
23.06.2022	23.4	750.4	754.6	-0.1	95	从东北偏东	1				间歇性雨，在整个相时在整体相时	23.0	26.0				2500或更高，或无云。	9.0	22.6	10	12			
23.06.2022	23.2	750.5	754.7	0.4	93	从北方吹来	1		4		间歇性雨，小雨，在观测时较弱	22.9	26.0				2500或更高，或无云。	3.6	22.0	3	12			
23.06.2022	23.4	750.1	754.3	-0.5	92	从东北偏东	1				间歇性雨，不冻，在观测时较弱	22.9	26.0				2500或更高，或无云。	10.0	22.0	无降水	7	12		
23.06.2022	24.1	750.6	754.8	-0.7	90	从东北偏东	1				云量发展并在整个相时在整体相时内云遮住了一半或以下的天空	22.9	26.0				2500或更高，或无云。	20.0	22.4	7	12			
22.06.2022	24.2	751.3	755.5	0.4	90	从东北偏东	1				小雨（不冻）	22.9	26.0				2500或更高，或无云。	30.0	22.5	30	12			
22.06.2022	24.1	750.9	755.1	1.1	92	从东北偏东	1		4		小雨（不冻）	22.9	26.0				2500或更高，或无云。	26.0	22.7	37	12			
22.06.2022	24.6	749.8	754.0	-0.6	96	从东方吹来	2		7		小雨（不冻）在整个相时在整体相时	22.9	26.0				2500或更高，或无云。	3.6	23.9	78	12			
22.06.2022	24.3	750.4	754.6	-0.7	96	从北方吹来	1		7		间歇性雨，在整个相时在整体相时	22.9	26.0				2500或更高，或无云。	1.4	23.6	72	12			
22.06.2022	25.5	751.1	755.3	-0.3	91	从东北偏东	1		7		小雨（不冻）在整个相时在整体相时	22.9	26.0				2500或更高，或无云。	27.0	23.9	63	12			
22.06.2022	23.1	751.4	755.5	1.6	97	从北方吹来	2		7		间歇性雨，在整个相时在整体相时	23.0	34.6				2500或更高，或无云。	0.4	22.6	57	12			
22.06.2022	24.3	749.6	754.0	-0.1	95	从东北偏东	1		7		间歇性雨，不冻，在观测时较弱	23.4	34.6				2500或更高，或无云。	8.0	23.4	15	12			
22.06.2022	23.5	749.9	754.0	-0.8	95	从东北偏东	1		9		小雨（不冻）在整个相时在整体相时	23.4	34.6				2500或更高，或无云。	16.0	22.7	15	12			
21.06.2022	25.3	750.7	754.9	1.5	82	从东北偏东	2		10		间歇性雨，不冻，在观测时较弱	25.3	34.6				2500或更高，或无云。	30.0	22.0	0.1	12			
21.06.2022	30.9	749.2	753.4	1.2	71	从西北偏西	1			11	云量发展并在整个相时在整体相时	26.8	34.6				2500或更高，或无云。	30.0	25.0	无降水	12			
21.06.2022	33.2	748.0	752.2	-1.1	58	从西南偏西	4			11	云量发展并在整个相时在整体相时	26.8	34.6				2500或更高，或无云。	30.0	23.8	无降水	12			
21.06.2022	32.8	749.1	753.2	-1.3	60	从西南偏西	4				云量发展并在整个相时在整体相时	26.8	33.3				2500或更高，或无云。	30.0	24.0	无降水	12			
21.06.2022	31.6	750.4	754.5	-0.3	65	从西南偏西	5				云量发展并在整个相时在整体相时	26.8	32.3				2500或更高，或无云。	30.0	24.2	无降水	12			

图一 天气数据

3 数据分析

3.1 数据预处理

- 1) pandas读取excel表格，分析图表内数据结构
- 2) 时间索引使用datetime函数重新排列
- 3) 分析数据间相关性，决定哪些列对于数据影响小以此进行数据集的填充和划分
- 4) 去除无关列如”ff10”等列
- 5) 字符类型转换为数值类型，可以使用独热编码
- 6) 缺失值填充，使用平均、中位数、相关性进行填充，由于部分数据具有前后的时序关系，因此基于这种时序关系的填充也是可行的。

3.2 相关性分析

绘制了数据特征间的相关性热力图（图2）

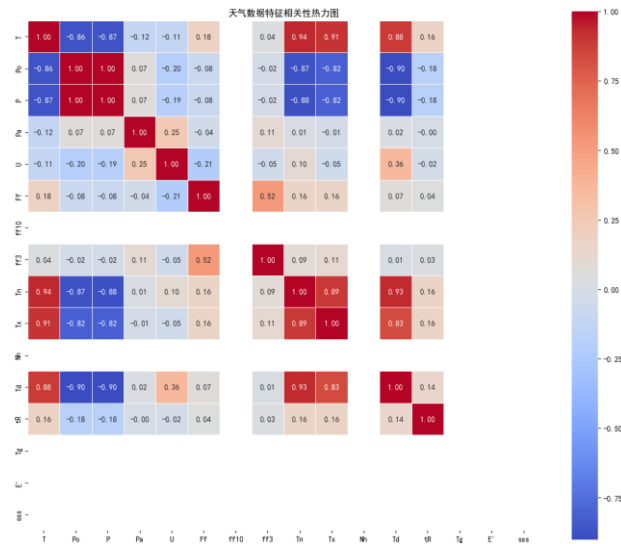


图2 相关性热力图

空白列代表整列缺失的数据，因此有需要对数据中毫无作用的空白列和对结果影响较小的列删除，在删除后的列的热力图如下图3：

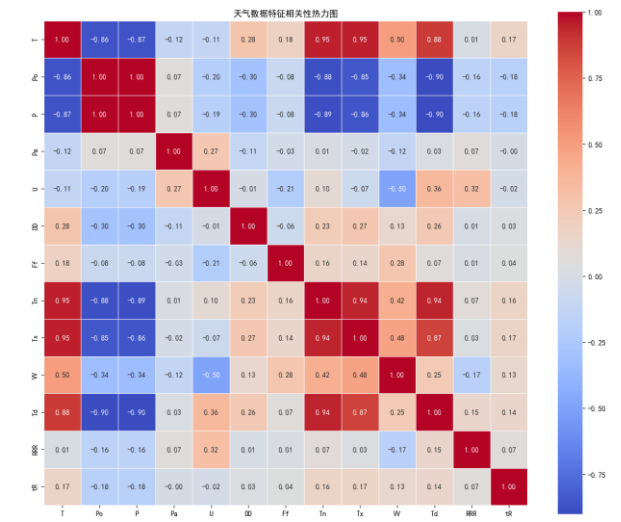


图3 处理后的相关性热力图

可以发现，天气和气压之间的关系呈现出明显的负相关，预测也大致是通过这两部分的数据进行预测。后续对这部分数据进行缺失值填补。

3.3 缺失值填补

天气的缺失值往往出现在一天内的某几次抽样，但实际上一天内的其他时刻的天气数据仍然存在。通常情况下，一天内的天气不会出现较大的突变，因此使用前后数个采样点的温度进行填充是可靠的。对于单个样本点的填补，使用均值填充；而对于连续数个样本点的缺失，遵循着一天内14点气温最高，5点最低的原则，进行带趋势插值填补。

而湿度、气压这类温度往往关乎于实际的各项指标，此时需要考虑和其他数据的相关性，彼此互相填充缺失的值。

4 模型训练和预测

4.1 分类模型

天气需要使用分类模型进行多分类预测，通过训练BP神经网络、XGBoost模型和极端树模型进行分类。使用气温、气压、能见度等列的特征进行训练。

训练需要将数据预先进行标准化，并转换为torch张量，调用dataloader的api构建数据迭代器，考虑到传统机器学习模型不支持torch库的api，保留了numpy格式的数据集进行调用。

BP神经网络使用Adam和交叉熵损失，最后均对结果进行accuracy评估，结果如下：

BPNN Accuracy: 0.6880356530682208

XGBoost Accuracy: 0.7123757284881728

ExtraTrees Accuracy: 0.7271169009256085

4.2 回归模型

对气温和湿度进行回归预测，气温具有明显的时序特征，因此可以使用LSTM模型进行训练，考虑到天气特征的独特性和大量数据带来的长程依赖问题，使用transformer模型进行比较，还训练了一个SVM模型进行预测。

湿度特征有明显的数据集中性，随着气温上升，湿度的比重可能会发生变化，由此可以认为湿度同样具有一定的时效性和先后性，同样可以使用LSTM进行预测。

气温预测的LSTM和transformer模型使用Adam和MSE，使用百分误差分析相对温度的偏

差量，并对未来数天的天气进行预测。结果如下：

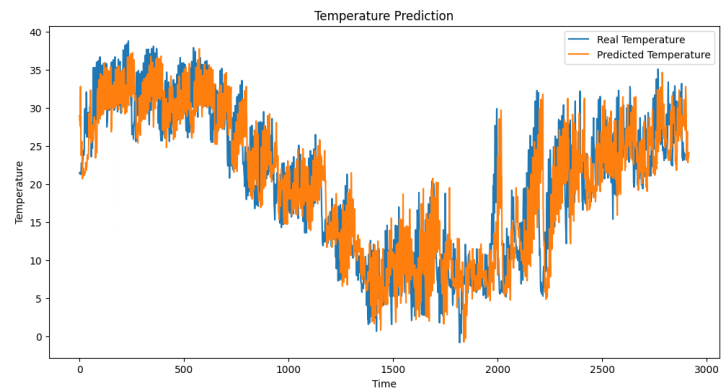


图4 LSTM的气温预测曲线

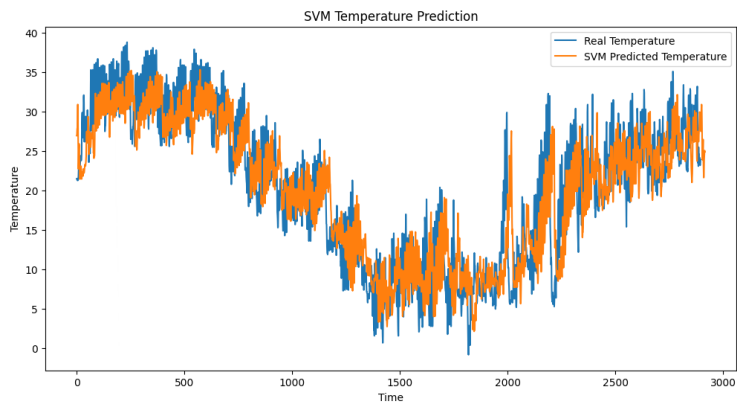


图5 SVM的气温预测曲线

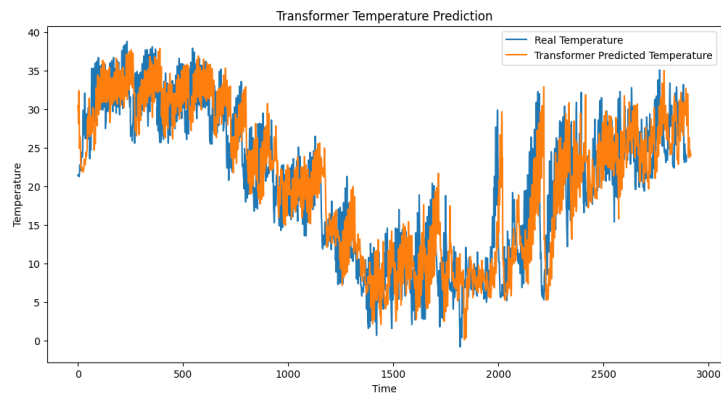


图6 Transformer的气温预测曲线

气温百分误差:

Transformer Relative Percentage Error (RPE): 35.24%

SVM Relative Percentage Error (RPE): 34.25%

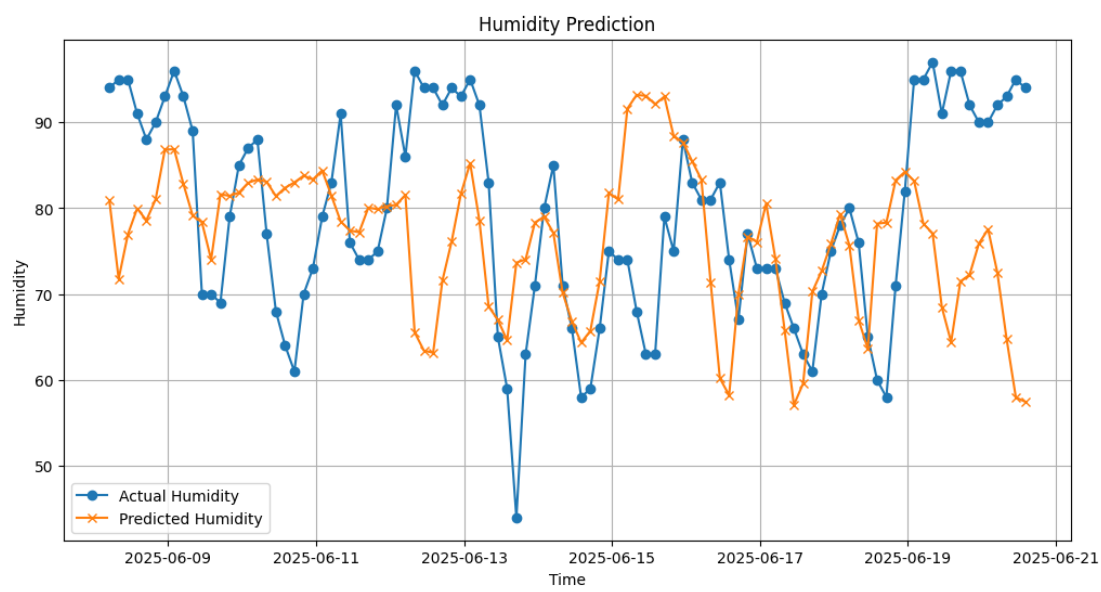
LSTM Relative Percentage Error (RPE): 35.37%

数天天气预测:

LSTM未来7天预测气温: [24.10558803 23.7061928 23.1464587 22.76283808
22.62319754 22.84877391 23.39280601]

Transformer未来7天气温预测: [24.80605165 25.56591378 25.96563829 26.12505932
26.11812876 26.13701991 26.24954493]

湿度模型训练结果如下:



5 结论

在本研究中，我们对湿度和温度的预测模型进行了深入分析，使用了Transformer、SVM和LSTM三种不同的机器学习模型。实验结果表明，这些模型在一定程度上能够捕捉到湿度和温度的变化趋势，但仍然存在一定的误差。

具体而言，湿度预测中，实际湿度与预测湿度之间的波动较为一致，表明模型在捕捉湿度变化趋势方面具有一定的能力。然而，温度预测的相对百分比误差（RPE）显示，Transformer、SVM和LSTM模型的误差分别为35.24%、34.25%和35.37%。这些误差表明模型在温度预测中的精度仍有待提高。

此外，未来7天的气温预测结果显示，LSTM和Transformer模型的预测值与实际值之间存在一定差异。尽管这些模型能够大致预测出温度的变化趋势，但在具体数值上仍有较大偏差。

综上所述，尽管所提出的模型在湿度和温度预测中表现出一定的潜力，但其预测精度尚不足以满足实际应用的需求。未来研究可以考虑引入更多的特征变量、优化模型参数或探索更先进的算法，以进一步提高模型的预测性能。此外，数据的质量和数量也可能影响模型的表现，因此在数据收集和预处理阶段也需要更加谨慎和细致。

6 参考文献

- [1] Lorenc A C. Analysis methods for numerical weather prediction[J]. Quarterly Journal of the Royal Meteorological Society, 1986, 112(474): 1177-1194.
- [2] Singh N, Chaturvedi S, Akhter S. Weather forecasting using machine learning algorithm[C]//2019 International Conference on Signal Processing and Communication (ICSC). IEEE, 2019: 171-174.
- [3] 黄天文, 焦飞, 伍志方. 一种基于迁移学习和长短期记忆神经网络的降水预报方法[J]. 暴雨灾害, 2024, 43(1): 45-53.
- [4] 孟祥福, 石皓源. 基于 Transformer 模型的时序数据预测方法综述[J]. Journal of Frontiers of Computer Science & Technology, 2025, 19(1).
- [5] 向德萍, 张普, 向世明, 等. 基于 Transformer 的多模态气象预测[J]. Journal of Computer Engineering & Applications, 2023, 59(10).
- [6] Willard J D, Harrington P, Subramanian S, et al. Analyzing and exploring training recipes for large-scale transformer-based weather prediction[J]. Artificial Intelligence for the Earth Systems, 2025, 4(2): 240061.
- [7] Karevan Z, Suykens J A K. Transductive LSTM for time-series prediction: An application to weather forecasting[J]. Neural Networks, 2020, 125: 1-9.
- [8] Srivastava A, Anto S. Weather prediction using lstm neural networks[C]//2022 IEEE 7th International conference for Convergence in Technology (I2CT). IEEE, 2022: 1-4.
- [9] Biswas M, Dhoom T, Barua S. Weather forecast prediction: an integrated approach for analyzing and measuring weather data[J]. International Journal of Computer Applications, 2018, 182(34): 20-24.
- [10] Sheikh F, Karthick S, Malathi D, et al. Analysis of data mining techniques for weather prediction[J]. Indian Journal of Science and Technology, 2016, 9(38): 1-9.