****

**电气与电气工程学院**

**2022-2023学年第一学期**

**招商证券软件工程训练营**

**综合作业报告**

**新能源多时空尺度出力预测**

|  |  |  |
| --- | --- | --- |
| 班级 | 学号 | 姓名 |
| 电气2009班 | U202012478 |  |
| 电气2009班 | U202012465 |  |
| 电气2009班 | U202012476 |  |
| 电气2007班 | U202012397 |  |

华中科技大学电气与电子工程学院

2022年09月

**1 研究背景和项目目标**

1.1选题依据

实现“双碳”目标，能源是主战场、电力是主力军。作为我国碳排放占比最大的单一行业，电力行业减排将直接影响碳达峰、碳中和的整体进程。构建新型电力系统，是实现“双碳”目标的重要支撑，而构建以新能源为主体的新型电力系统迫切需要从全网统筹的角度对新能源功率进行更加精确化的预测，所以我们选择了本题目，并希望用深度学习算法来建立预测综合能力更强的模型，为推进新型电力系统的建设和为我国“双碳”目标下构建清洁低碳安全高效能源体系贡献华科电气人的思考和努力。

该题目需要解决问题是：1.在我国，区域地形地貌丰富、气候复杂多变、新能源场站点多面广、基础条件参差不齐，传统以单一算法为技术路线的新能源功率预测方法难以适应复杂多维场景下的精确预测，需要通过“开放式”预测测试解决多主体来源的先进预测算法灵活应用问题。2.新能源预测模型适应性差、预测结果可信度不高、极端天气预测“失灵”等关键共性问题，也是制约当下新能源预测能力提升的主要障碍，亟需从技术和管理上进行创新突破。

1.2业界现状介绍

自2019年以来，不断有新的新能源多时空尺度发电预测平台建成投运，新能源多时空尺度精确预测技术招标公告也时有出现，各方论坛也正聚焦于这一技术。2022年新型电力系统调度运行和电力市场创新技术论坛在广州举办、新能源多时空尺度精确预测技术研究南方电网招标、新能源多时空尺度出力预测公共算例的比赛等可见，在全面推进构建新型电力系统的趋势下，新能源多时空尺度预测技术正成为行业热点，也成为了我们现在急需解决的问题。

业界现有预测技术中，一般通过采用趋势外推、线性回归及智能预测方法（如神经网络）等传统预测方法，一方面由于没有建立全面的影响因素，另一方面由于样本量较少而使得预测方法的预测性能不能完全体现，导致预测精度不高，因此，目前新能源出力预测方法不能满足对预测精度的要求，这将显著影响到电网调度的安全性和经济性。

总结预测模型问题所在：一是传统单一算法存在缺点，二是新能源预测模型适应性、可信度等关键共性问题有待突破。虽然已有发明提供了一种新能源出力预测方法、系统、存储介质及设备，引入多层模型预测用于解决现有技术中的新能源出力预测方法不能满足对预测精度的要求，但其现实实现及实际效果仍待考究。

1.3 本项目的目标

本项目目标如下：1.对数据进行清理，并开展探索性数据分析，并给出初步的分析；2.根据分析结果抽象数学模型，并采用机器学习与深度神经网络等方法建立初步模型，对未来7日的新能源出力进行预测，给出结果；3.分析预测结果并根据结果对模型做出改进 4.最终基于模型做出一个预测软件，能够根据用户输入数据给出预测。5.终极目标：做出一个自主训练更新模型的软件，能够根据输入数据自动分析模型预测效果进行更新。

**2 项目总体设计**

2.1工程问题分解

对于新能源多时空尺度出力预测项目而言，这本身就是对一个大型的工程问题设计一个合适的方案，那么从工程本身考虑，就需要我们建立合适的数学模型作为预测基础，然后设计一个基于此模型的软件来做工程应用。而从项目业界情况来看，核心的问题在于算法与模型上的问题，所以项目关键问题也就转移到了使用先进的算法来建立更优化的模型，弥补传统预测模型和方法上的不足。由上可知我们要解决的核心问题便分解如下：a.如何使用深度学习等更先进算法提高预测力，b.怎样应用算法建立优化且合适的预测模型；c.在解决a,b问题基础上做出相关软件实现工程应用。

2.2项目方案设计

从三个问题便可对整个项目进行设计，从中要解决a，b问题就需要算法技术解决和创新，这个算法模块其实是和建模是一体的，且不可分割，所以就需要三个模块，第一个模块实现对已经有结果的数据进行训练和预测，得出相关的预测评估，并能选择出合适算法，从而实现对数据较好的预测（基本设计思路是使用LSTM和多层深度学习结合实现）；第二个模块就是基于已选算法应用到实际预测中，得出预测结果，把预测做成一个界面，从而实现训练和预测一体化功能。第三个模块则是解决问题c,项目团队将做出一个软件，并尽量让软件能够实现自主训练更新，通过选定的训练集和预测集训练实现多方位的预测，实现工程实际应用价值。

2.3设计评估

设计最复杂与核心的点是模型建立中的算法确定，因为只有采用合适的优化算法才能得出最佳的模型，解决之前模型难以精准预测以及适应性差，可信度不够高等问题。此点设计难度高，且由于项目团队本身获得的数据有限，所以选择和评估难度加大，但是采用深度学习后，模型的拟合结果还是比较让人满意的。另外项目中一个技术难点就是如何做出真正完整的预测软件，虽然没能做出真正的软件，因为tensorflow库与软件形成函数有冲突，但是在团队成员的努力下，设计的交互界面的功能也算完善，能够完成预测功能，并增加了查找表的功能，能对单点数据进行查找相似数据点得出大概的预测值，也能够选择预测文件和训练的文件并更改训练次数以及训练所记忆的时间长短，并导出结果和绘制预测图像。

2.4 项目组织实施

整个项目的组织实施将按模块化分工，每个项目成员主负责一个模块，而对于技术要求与设计难度较大的模块，大家又会相互合作，共同突破难关，实现相关模块，总之是在一种分工+合作的模块化多方位推进下组织自己的解决方案，实施完成我们的项目。

**3 项目关键技术**

3.1基本点介绍

项目建模与软件实现过程中使用的语言是Python,运用的工具是Visual Studio Code 和pycharm，导入的库包括tensorflow (使用机器学习、神经网络的库)、pandas(用来导入数据的库)、numpy（用于做数值计算扩展，存储和处理大型矩阵的库） 、sklearn.preprocessing（做机器学习预处理的库，我们导入了库中的MinMaxScaler来做归一化）、matplotlib（数据可视化库，导入了其中的pyplot用来做数据可视化）、seaborn （绘图的库）、PyQT5(跨平台的 C++ 开发库，主要用来开发图形用户界面（Graphical User Interface，[GUI](https://so.csdn.net/so/search?q=GUI&spm=1001.2101.3001.7020" \t "_blank)）程序，调用了里面的类，如QtCore:包含非核心的GUI功能，此模块用于处理时间、文件和目录、各种数据类型、流、URL、MIME类型、线程或进程；如QtGui：包括窗口系统集成、事件处理、二维图形、基本成像、字体和文本)

3.2算法：基于LSTM的优化深度学习算法

长短期记忆网络（[LSTM](https://baike.baidu.com/item/LSTM/17541102" \t "_blank)，Long Short-Term Memory）是一种时间循环神经网络,也是我们项目建模的核心算法，我们调用tensorflow实现使用LSTM结构的循环神经网络来进行数据处理与预测，而训练时，为了最小训练化训练误差，我们使用了Adam算法（adaptive moment estimation）。我们在数据处理时首先调用函数进行了相关性分析，之后对数据进行了归一化，并且对相关性拟合较差的数据使用多层的LSTM网络进行多次预测，可以进行多达几十次的深度预测，以求得到较好的预测结果（基本思路即是将LSTM神经网络预测结果值作为输入的新特征量进行再次预测，再次预测后得到结果再与之前的特征量整合到一起，我们再次输入进行预测，进行深度多层预测，最后得到预测值）所以我们的核心技术也就是使用基于LSTM进行优化建立的深度学习算法。

3.3模型

本项目主要采用双层的LSTM模型对数据进行处理、分析、预测，并以双层LSTM模型为基础，进行深度的模型建立与预测，以求得到较好的结果，主要过程是将预测得到的结果作为新的输入特征量，再用双层LSTM模型对此特征量进行训练并做出相应的预测，以求得到相关性更高的特征量来得到更加趋于拟合的结果。

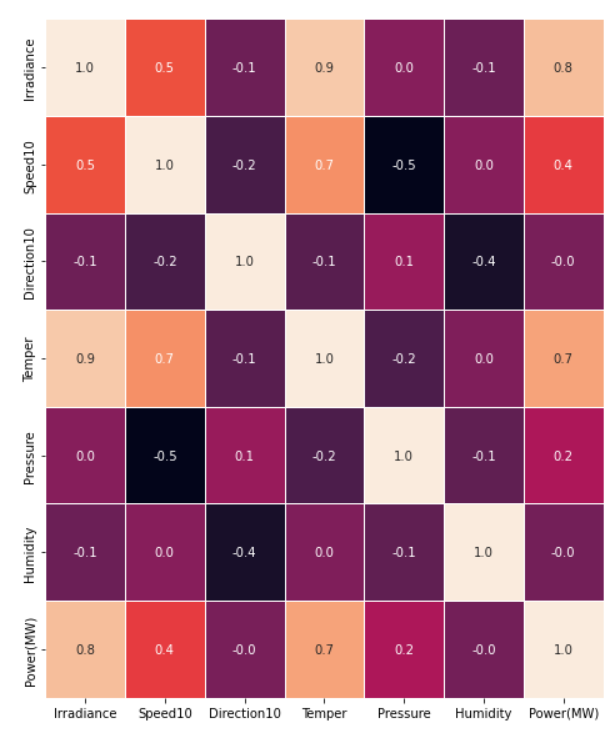
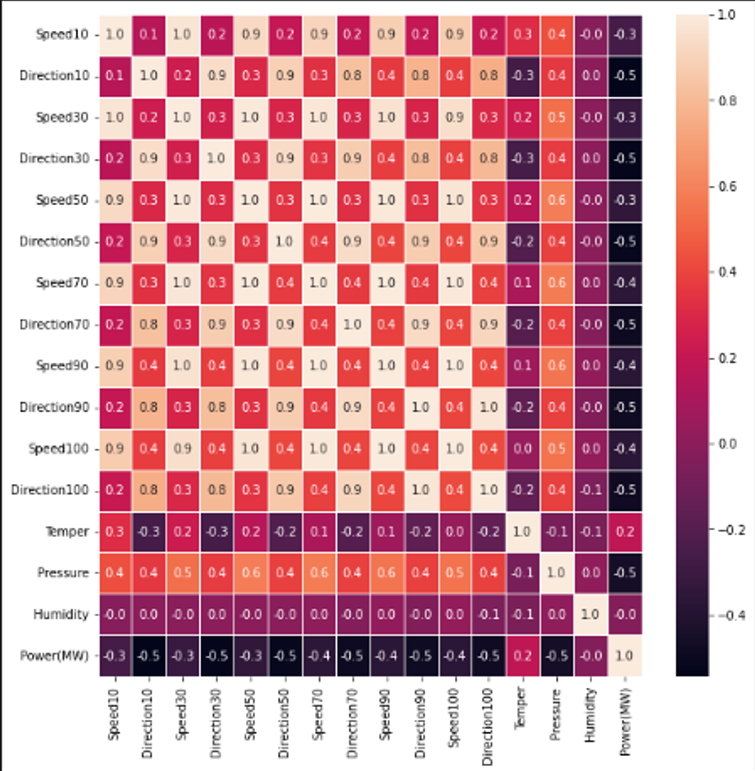
3.4软件实现

使用PyQt5工具包搭建用户界面，再通过函数链接实现模型训练、预测和数据文件读写等功能，做出一个由用户进行选择文件或输入数据就可以实施相应能源数据（风电、火电）预测的软件。

**4 项目实现**

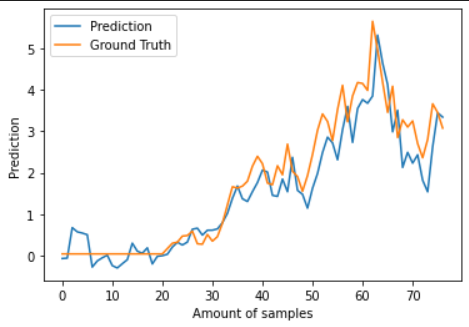
项目具体实现过程中，在大的方向上其实仍是数据预测算法建模和软件链接两大块，分别将两大模块链接统一实现起来就完成了我们整个新能源多功能尺度出力预测的工程实现。

在用基于LSTM的优化深度学习算法前，我们首先对所给的数据进行处理得到了相关性热力图，以便于后续方便根据相关性参数进行预测，图片如图1。



**图1 风电以及光电的结果与各特征量的热力相关图**

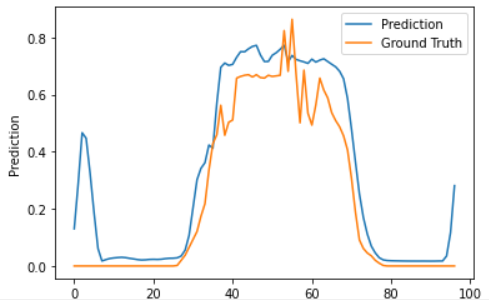
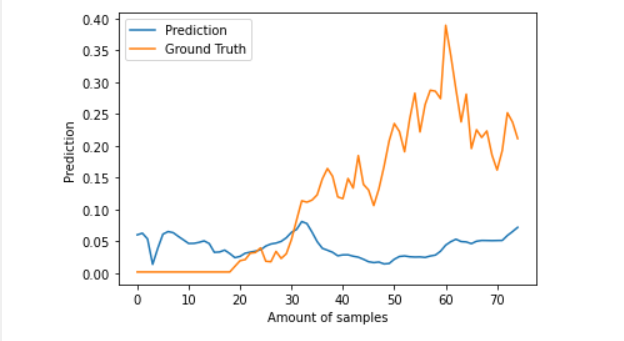
从图1中我们可以发现风电参数的相关性较差，而光伏发电的相关性有几个参数较好，由于风电的相关性较差，故需要采用很深层的网络进行预测，可能最终的效果也不会很好，故我们对于光伏的分析要更多一些。首先再进行模型的训练之前，我们采用归一化将模型参数进行拟合。之后，由于最初没有看到给的预测所用的数据不包含结果，所以使用的是十六维的数据（即包含所给出的结果的数据）进行模型的训练，由于第十六维就是结果，相关性为1，很容易得到符合较好的函数：



**图2 16维风电预测**

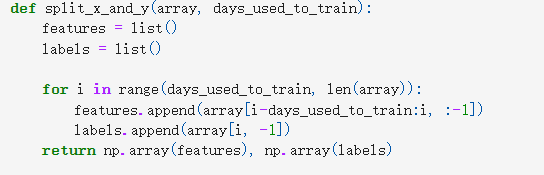
但是，当我们去除输出结果，对模型进行训练的时候，不难发现：模型的拟合程度非常差，甚至没有任何的拟合，可以这么说，总结来说，可能是数据量不够，或者模型的训练深度不够，或者是由于数据的相关性实在太差导致的，故我们打算采用深层的LSTM模型进行预测：

如若不采用深度预测，得到结果如下：



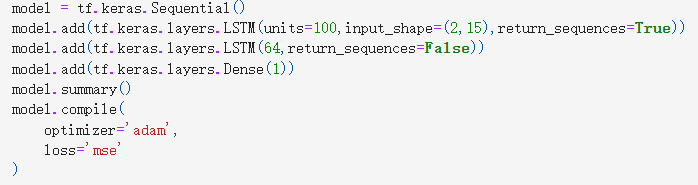
**图3 不进行深度预测-风电 图4 不进行深度预测-光伏**

首先新定义了一个划分数据块的函数，也是记忆时间的长度，这里定义的函数可以随意改变用于记忆的天数，可以用于模型的训练：



**图5 用于选取记忆时间的函数**

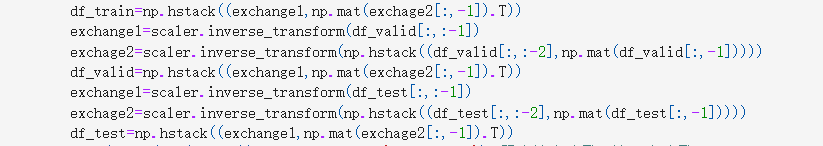
然后训练模型，这里采用两层的LSTM网络，使网络更深，第一层输出多维的数据，第二层为一维的数据，以便得到结果，当然也可以采用更加深层的模型，但可能会过拟合，而且本身数据就需要不断的训练，然后得到更加深度的模型。故最终只采用了两层LSTM，并加入Dropout层，防止训练深度大导致过拟合：



**图6 训练的神经网络层**

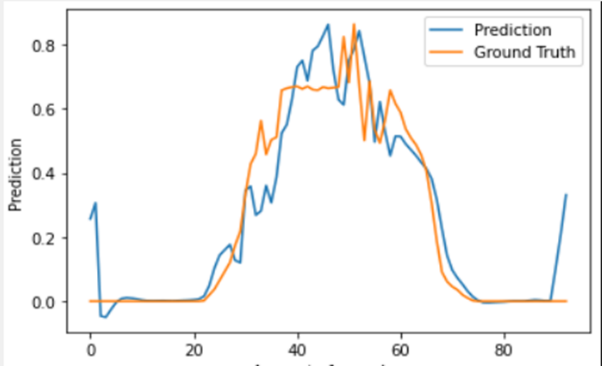
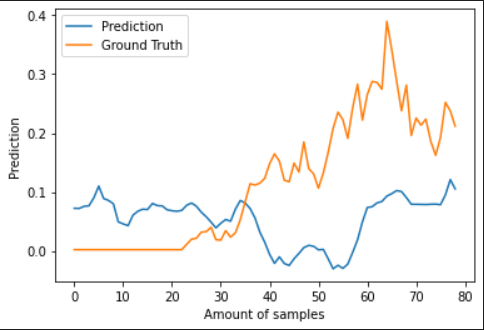
然后通过这十五维的数据，对结果进行预测，然后得到预测的结果，再把这个预测的结果加入最开始的数据中，作为第十六维的特征输入，并用这十六维的特征去进行预测，并一直继续下去，可以得到更为深度的模型以及更多的特征量。

然后对数据进行反归一化处理，将原本的数据的最初选取用于训练的行数取出，并与结果进行合并，用inverse\_transform函数即可得到反归一化的结果。

****

**图7 反归一化**

在这里算力有限，故只进行了两层模型的建立和预测，得到的结果如下：

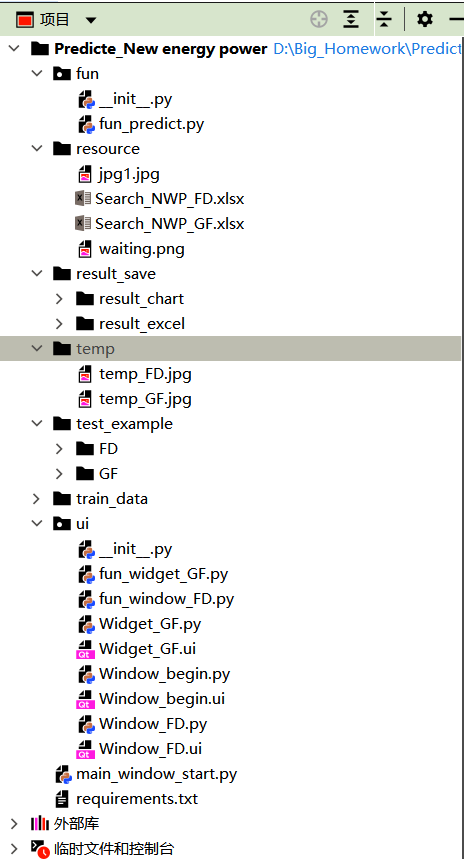
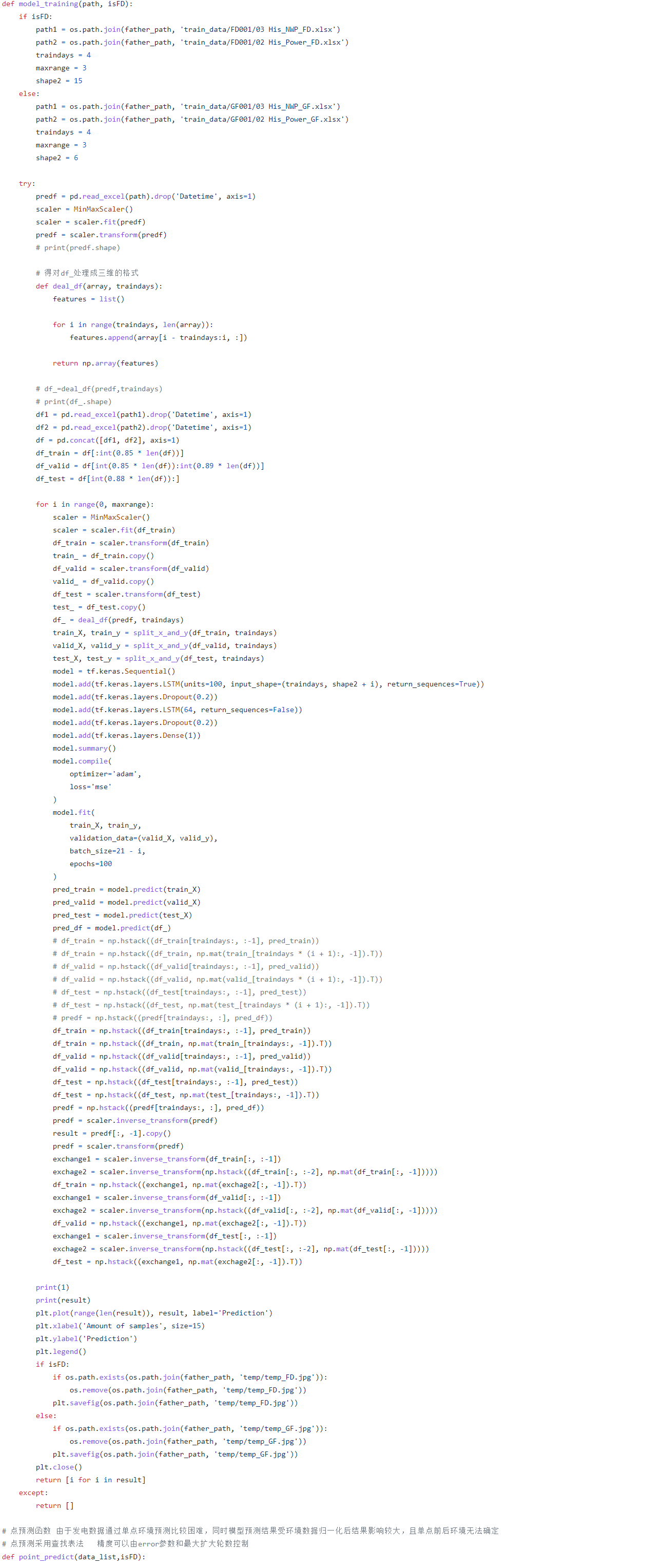
****

**图8 风电稍微深度预测的结果 图9 光伏深度的预测结果**

由此可见在数据预测模块，使用深度预测模型的话可以得到相比不进行深度预测更准确的预测结果，在中间的折线段，更加的陡峭而不是趋于平缓，而不进行深度预测结果则差强人意，从预测图前后的对比来看，我们可以知道预测准确度上，风电预测其实还是未达到精准要求，而光伏的深度预测已经可以用来做准确性预测。其实这里也反映出我们算法模型本身的缺点就是在于即使采用了深度预测也偏向适合相关性较强的预测，而对于参数相关性较弱的风电预测则有一点束手无策。另外，给出的数据不足肯定也是预测不能选出更好方案和更深层准确预测的一个原因，但是我们模型本身的预测性较强和拓展性强的优点也还是很明显的，并且对光伏预测处理也是得出了预期结果。

不过我们也能给出我们模型的进一步优化方案：那就是使用更深层的模型进行预测，而且只要电脑运算能力强，其实模型实施上可以进行层数更多维度的运算，得到更精确的结果。也可以适当的更改参数使模型与结果更加吻合，当然，这样的优化方案可能造成的不良后果就是存在过拟合的问题，但是如果在数据样本足够的情况，我们是可以依据实验平台数据得到准确拟合且效果更好的预测模型结果。

最后我们将模型拟合成了一个函数，以便更好地应用到GUI界面，然后进行大量的重复训练以得到最优的结果：



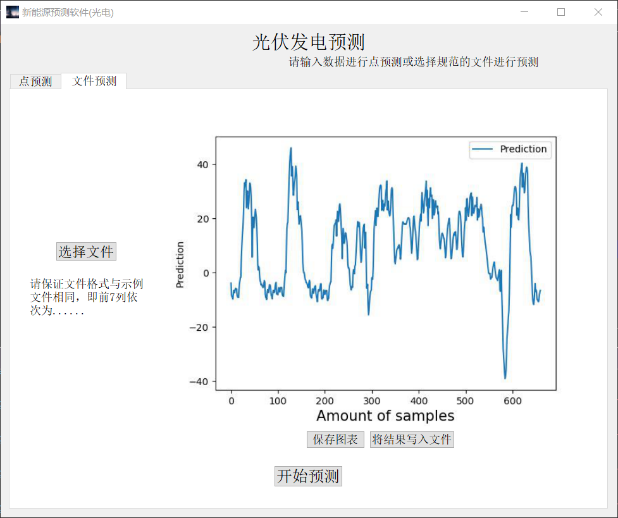
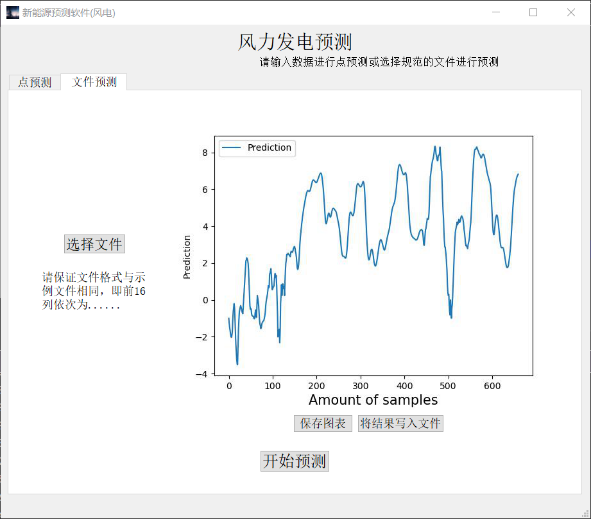
**图10 模型训练函数及项目目录包含函数**

软件前端这一块看似是项目应用实现和模型建立的延伸，但其实本身工作量也不小。

图10是交互式界面的项目目录，包含所需函数和文件。fun文件夹是实现预测功能的模块，fun\_predict.py文件中包含LSTM神经网络模型的训练和预测功能的函数，以及用于实现文件存取操作的函数，实现了模型、数据和软件界面的链接。resourse存放图标等资源文件；result\_save存放保存的预测结果；temp存放临时文件；test\_example存放待预测的环境参数文件；train\_data存放用于模型训练的数据。ui文件夹是与软件界面相关的模块。Widget\_GF.py、Window\_FD.py、Window\_begin.py分别是光伏预测、风电预测和主界面的配置文件，通过QtDesigner设计好界面的控件和显示，得到.ui文件，再转化为上述相应的.py文件。fun\_widget\_GF和fun\_window\_FD文件定义了软件界面的控件功能和显示内容，可以实现模型训练、预测，结果显示和文件选择、存取等操作。最后main\_window\_start.py是整个项目的程序入口，运行该文件即可启动软件。完整的预测软件就由上述文件组成的。虽然整个项目实现过程中将数据预测模型和软件界面连接还是比较困难，但整个项目团队还是成功了将数据文件导入链接起来，通过软件界面和数据处理模块的配合得到了一个完整的软件，展示如下：



**图11 主界面**



**图12 风电预测界面及预测结果 图13 光伏预测界面及预测结果**

从图11、12、13我们可以发现整个项目已经实现了制作出一个软件平台，整个项目体系也彻底实现，基本完成了最先设定的项目目标，只要对整个项目进行打包即可生成具有相应功能的exe文件。项目中的一些地方还可以进行后续优化和调整，使性能得到提高，软件更加轻量，从而得到更实用的新能源发电预测软件。

**5 项目测试**

通过整个项目的实施和推荐，我们已经能够做出一个软件预测，能根据用户给出的文件数据或输入的点数据预测实现风电和光伏新能源出力预测，在功能实现了能够根据数据需求进行预测并得到较准确结果的目的，在性能上虽然可能模型仍达到高精确性和适应性的目的，并且应用仍未完善到高级目标，但模型和软件性能已能作为一个工程应用软件。对整个项目测试后，我们还是基本完成了最低设计的目标，实现了由算法选择-数据处理-模型工程应用的项目完成，虽然因为数据集较少以及高难度核心技术为突破未能达到做出一款自我训练优化模型的软件，但项目还是基本实现了原有价值，并且项目拓展性也较强，后期可以继续开发与技术攻关，测试得到一款全新的新能源多尺度出力预测工程软件。

**6 项目管理**

6.1 团队人员组成

李亮霖、黄炼、杨文博、刘子贺。

6.2 任务分工

李亮霖：负责模型分析，模型建立，数据处理。李亮霖对于整个项目算法与数据处理上起到了一个核心实现作用，在整个团队中完成了题目要求部分并推进了项目的实施。（25%）

黄炼：负责数据处理，软件搭建，模型应用。黄炼在整个团队中主要负责软件搭建与模型应用，将整个项目实现了由模型到工程的推荐，让整个项目的工程意义更深刻。（25%）

杨文博：负责数据处理与模型分析，报告撰写，答辩与模型拓展。杨文博在项目负责了初步模型的分析，提供了初步模型思路，并将项目成果撰写与报告中并进行答辩，从在后期工程角度拓展模型。（25%）

刘子贺：负责模型建立，PPT制作，模型建立后模型优化。刘子贺辅助了模型的建立，并制作了PPT来汇报整个项目的完成历程，并在模型建立后分析模型的优点不足对模型进行一定程度的优化。（25%）

贡献比分配（25%，25%，25%，25%）

**7. 总结与反思**

在本次项目中，我们完成了题目本身的要求，实现了新能源多时空尺度模型的建立并做出了工程预测相关界面，基本实现了用自己的深度学习算法模型来解决工程问题的目标，通过自己的项目做出了作为华科电气人对实现双碳目标下新型电力系统新能源预测的独特贡献。

在本项目中，我们做的好的地方首先就是能够合理利用课程知识，将课程算法与培养的工程意识应用到项目中，做到了知行合一，其次就是我们项目始终有合理的分工和交流，大家各自负责项目的一个主攻部分，为项目推进贡献了自己的独特力量，但又相互配合，共同解决技术和代码上的问题，在交流中改进和实现了项目。而且我们项目设计目标也梯度也较合理，让大家既能够脚踏实地做出部分成果，又充分考虑了项目难度和工程应用，能够在未来继续拓展项目。至于项目中不足的还是由于深度学习和人工智能知识上的缺乏和实际软件工程项目实践经验的欠缺，导致项目并没有完成做出一个成熟的软件产品，并且真正的能实现现有技术上的很高突破，不够项目还是具有未来计划和实施空间的。

总而言之，整个团队在整个项目中不仅将所学知识在实际工程背景下应用，锻炼并培养了自己的工程意识，还在从代码编写，程序实现，报告与PPT展示等整个一体的项目中相互交流与成长，培养了团队意识，提升了合作能力，为以后成为电气工程人打下了更好的基础，收获颇丰。

附件4 课程项目设计报告评分表

|  |  |  |  |
| --- | --- | --- | --- |
| **过程要素** | **评分依据** | **分值** | **得分** |
| 1、项目具有明确的工程和社会意义 (6分) | （1）表达出项目具有明确的实际意义，同时目标明确且合理，项目的实施能运用到课程学到的知识 | 6 |  |
| （2）无法表达出项目具有明确的实际意义，同时目标明确但不太合理，运用了部分可能学到的知识 | 4 |
| （3）无法表达出项目具有明确的实际意义，没有明确的项目不表，没有运用了学到的知识 | 2 |
| （4）无表述 | 0 |
| 2、总体设计合理，体现工程思想 (6分) | （1）总体设计合理，体现工程思想，问题分解粒度合适，整体各个模块项目组织合理 | 6 |  |
| （2）总体设计较为合理，体现部分工程思想，问题分解粒度不合适，整体各个模块有一定的组织 | 4 |
| （3）总体设计随意，没有体现工程思想，问题分解不合理，整体各个模块项目组织合理 | 2 |
| （4）没有表述 | 0 |
| 3、项目技术实现(6分) | （1）选择的运用的关键技术合理，项目实现工程量合理，有一定的技术挑战性，基本达到设计目标 | 6 |  |
| （2）选择的运用的关键技术较为合理，项目实现工程量较少，没有技术挑战性，基本达到设计目标 | 4 |
| （3）选择的运用的关键技术不合理，项目实现工程量极少，没有技术挑战性，基本没有达到设计目标 | 2 |
| （4）没有表述 | 0 |
| 4、项目管理水平(6分) | （1）团队建设合理，成员分工合理，共享较为均衡，时间进度按计划执行 | 6 |  |
| （2）团队建设较为合理，成员分工较为合理，共享不均衡，时间进度按计划执行 | 4 |
| （3）团队建设较为合理，成员分工不合理，共享不均衡，没有时间进度按计划执行 | 2 |
| （4）没有表述 | 0 |
| 5、从项目实施过程中能够总结出收获和教训(6分) | （1）明确知道本项目实施过程中，做的好的部分和做得不好的部分，有明确的未来改进的计划，能够总结课程学到的知识 | 6 |  |
| （2）基本知道本项目实施过程中，做的好的部分和做得不好的部分，有一定的未来改进的计划，能够总结部分课程学到的知识 | 4 |
| （3）基本知道本项目实施过程中，做的好的与坏的地方描述不准确，没有未来改进的计划，不能够总结部分课程学到的知识 | 2 |
| （4）没有表述 | 0 |
| 总分 |  |  |  |
| 评 语  批阅签名： | | | |

附件5 课程项目设计答辩评分表

|  |  |  |  |
| --- | --- | --- | --- |
| **组名** | **维度** | **具体要求** | **分数** |
|  | 学习氛围（10分） | 学习氛围浓厚，具有很强的团队凝聚力和团队协作能力(10分)。 |  |
| 学习氛围较为浓厚，团队能够主动协作开展项目研究工作(8分)。 |  |
| 团队能够合作完成项目研究工作(6分)。 |  |
| 设计、实验方案选择（10分） | 能够完成使用算法与模型的设计，对设计方案进行优选，模型与超参数选择正确合理，设计方案具有创新性（10分）。 |  |
| 能够完成算法与模型的设计，对设计方案进行优选，模型与超参数选择正确合理（8分）。 |  |
| 能够完成算法与模型的设计，模型与超参数选择基本正确（6分）。 |  |
| 分析与解决问题的能力（10分） | 能运用所学知识和技能去发现与解决实际问题；能正确处理实验数据，能对课题进行理论分析和科学计算与设计，并得出有价值的结论（10分）。 |  |
| 能运用所学知识和技能去发现与解决实际问题，能正确处理实验数据并得出正确结论（8分）。 |  |
| 能运用所学知识和技能去分析实际问题并得出结论（6分）。 |  |
| 成果展示情况（10分） | 研究报告撰写规范、论述清晰、结论明确；PPT制作认真，答辩陈述清晰、重点突出、回答问题正确（10分）。 |  |
| 研究报告撰写较为规范、论述比较清晰、结论有一定价值；PPT制作较为认真，答辩陈述比较清晰、回答问题基本正确（8分）。 |  |
| 研究报告撰写格式规范、论述较为清晰；PPT制作完整，答辩陈述较为清晰（6分）。 |  |