**电气工程学院2016级本科生毕业设计（论文）中期检查表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 学生姓名 | 叶星汝 | 学号 | 3160103013 | 专业 | 电气工程及其自动化 | |
| 论文题目 | 风电站预测与运营关键技术开发 | | | | 指导教师 | 王慧芳 |
| 项目介绍、原进度安排 | 项目介绍：  调研目前风电站预测方法的现状和最新进展，学习机器学习和人工智能算法和工具，利用人工智能、大数据挖掘、深度学习等智能技术对风电系统的实时数据和历史数据进行分析研究，实现基于大数据分析的风电数据的异常数据识别与修正技术，并将清洗后的数据应用于风电场短期预测在线建模中，实现对预测模型高效、高性能建模。讨论最新的人工智能技术用于风电预测的可行性，开发风电预测算法，并用算例进行验证。研究风电站运营平台的数据高效管理和数据库设计。  预期目标是优化风电预测算法、研发风电功率预测在线建模与优化系统。该系统基于高效计算框架，可以实现多类型/多层级的风电关联数据管理及快速读取/存储、大数据挖掘分析、预测特征优选与构建、深度学习预测模型的高效快速建模，并基于此发表一篇论文。  原进度安排：  2020年1-2月：学习Python、JavaScript算法知识、深度学习算法知识、风电功率预测知识  2020年3月：研究、优化分布式随机梯度下降算法  2020年4月：改进短期/超短期功率预测算法  2020年5月：结果整合与展示、形成论文成果 | | | | | |
| 目前  已完成  任务 | 学习JavaScript算法知识、学习机器学习算法知识，利用机器学习库Tensorflow.js完成模型搭建工作，学习如何在浏览器/node.js中进行机器学习。  机器学习的具体过程：   1. 选择数据(Select Data): 整合数据，将数据规范化成一个数据集，收集起来。 2. 数据预处理（Preprocess Data）: 数据格式化，数据清理，采样等。 3. 数据转换（Transform Data）: 这个阶段做特征工程。特征工程在机器学习中占有相当重要的地位。机器学习的上限取决于数据和特征，模型和算法只是逼近这个上限的方法。 4. 数据建模（Model Data）: 建立模型，评估模型并逐步优化。   依照上述机器学习的具体过程步骤完成的任务：   1. 选择数据：我的数据来源是2012年全球能源预测大赛的风能预测部分（Global Energy Forecasting Competition 2012 - Wind Forecasting）。它提供了一个数据集，其中包含这些风电场的历史功率测量数据，以及这些风电场级别的风力成分（纬向和经向分量，相应的风速和风向）的气象预报。 2. 数据预处理：由于这7个风电场的风电功率已经作好了归一化处理，我从已作好处理的csv文件中提取有效数据，进行数据洗牌（shuffle）和批处理（batch划分）。 3. 特征工程：2012年全球能源预测大赛的风能预测部分提供的数据提供了每个风电场每隔12小时发布未来48小时的天气预报，包括纬向和经向分量、风速、风向，这样做是为了模拟实际环境中的情况。但还缺少一些信息，如温度和空气密度，不过这些信息可以由时间、季节进行大致上的特征提取。（下一栏会阐述我在实际应用中对特征工程的困惑。） 4. 数据建模：我利用了Tensorflow.js在浏览器中建立模型并采用在实践中非常有效并且不需要配置的adam优化器。我的模型的评价指标为RMSE，即均方根误差，√[∑di^2/n]=Re。可视化评估部分：我目前只会使用最简单的tfvis-js可视化工具——散点图等。 | | | | | |
| 存在的  问题及拟采取的措施 | 1. 数据的读入要从异步改到同步，且希望能一次性读入所有数据而不是受制于toArray()的1000。 2. 大赛的题目是提前48小时预测，但我的预测可以不用局限于这个条件，因为按常理，时间越靠近，天气预报越准，对数据我暂时认为没有必要输入提前48小时的天气预报。 3. 对特征工程（降维）的理解不充分，比如在全球能源预测大赛的风能预测第一名的预测中，将天气预报中风的风速和风向（“ ws”和“ wd”）构建成特征wp~wd\_cut\*(ws+ws^2+ws^3)，他们给的理由是“对于此类特征处理方式，对于风力发电来说，不仅要考虑风力的强度，还要考虑风力的方向——风车在不同的方向上有不同的效率水平。风能转换与常数和ws(三次风强度)成正比， 它也取决于空气密度”。但我还没有完全理解。 | | | | | |
| 下一步 工作安排 | 2020年4月1日——2020年4月15日：  优化特征工程，结合时间序列模型知识、K均值聚类算法（k-means clustering algorithm）等知识，优化现有的的风电预测模型的算法。研究时间序列主要目的：根据已有的时间序列数据预测未来的变化。K均值聚类算法是一种简单的聚类分析学习算法，将n个实体最优划分为k组，使组内成员与其对应的代表组的质心之间的总距离最小。  利用机器学习可视化库tf-vis.js将风电原始数据、模型拟合过程和预测数据更加直观地体现出来。tfjs-vis是浏览器可视化中的一个小库，旨在与TensorFlow.js一起使用，它是一组用于可视化模型的工具，能可视化TensorFlow.js特定对象，并不会干扰我的Web应用程序。同时我还可以使用其他工具（例如d3，Chart.js或plotly.js），自定义可视化。  2020年4月16日——2020年4月30日  由于现有工作是基于2012年全球能源预测大赛风能预测部分提供的风电数据，随着风电技术的进步，数据不能很好地反映出当下的风电场环境，只能用于风电预测的训练。如果我能获取到更靠近当前时间的风电数据加以处理分析讨论会更加具有代表性与说服力。有余力的话，我会去找到更为复杂的海上风电数据进行风电功率预测，并同陆上风电功率预测进行比较与分析，讨论最新的人工智能技术用于海上风电预测的可行性与陆上风电预测的差异之处。 | | | | | |
| 指导教师  意见  和建议 | 指导教师（签名）：  2020年 月 日 | | | | | |

注：该《中期检查表》填写不少于2000字