回复

陈老师评论

一个是如何保证计算图可导，另一个是如何衡量概率分布的差异。比如我们用的是js，js相比于更常用的kl，具有对称性；带宽部分的设计，保证可导，pdf距离计算，这三点是重点；

实验的目的是为了衡量不同噪音的影响，而且是一种有偏噪音的影响。因为我们认为说这个有噪音的情况下，你的原始的真实概率分布的影响才会比较大，才能体现出它的优势。

主要贡献

第一次把概率分布加入到神经网络中，以前的知识融合只是把方程加入到神经网络，

第二是训练中的高噪声也是经常出现的，所i以实现场景的设计很有道理

其他建议：在浆距角加噪声；少量数据预测（两天预测半天）；噪音翻转

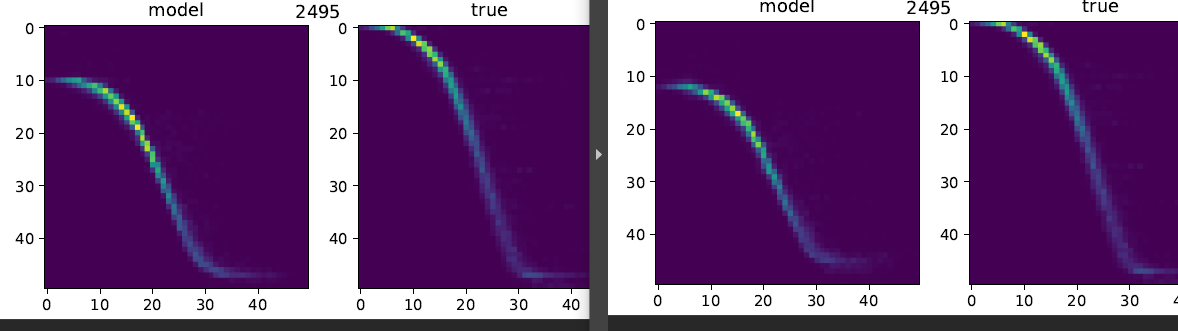
Gao Jiaxin at 28/6/2023

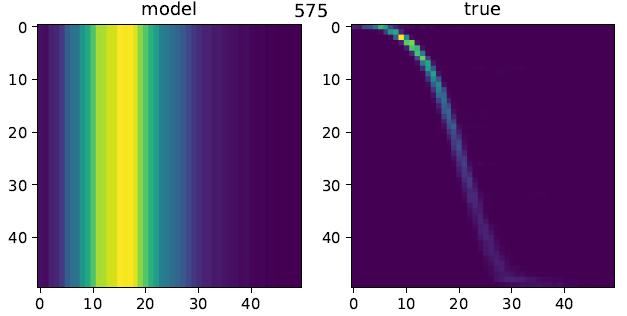
hello，之前的那篇论文陈老师给了很多修改意见，我现在正在改，有些地方要跟您再请教下：1）看您之前的报告中还提到了动态带宽，您还有动态带宽的相关记录吗，std为0的时候怎么处理呢？（陈老师还是说带宽这部分内容有点少，想多加一些，当一个创新点）2）陈老师想做一个纯js loss的消融实验，请问下纯js loss的时候如何调收敛？我自己试了很多随机种子或学习率，loss要么不降低，要么始终是0；3）在计算风功率曲线（包括真实的风功率曲线和根据模型预测结果得出的风功率曲线）的时候，此时风速有没有加对应的噪声？4）您之前说风功率曲线采用固定上下界会使模型的输出刚开始会集中分布在一个很窄的区间而无法求梯度，如果采用模型的预测值当上下界会出现这个问题吗？

上下界选择

Cheng Yuanqi at 11/4/2022 15:30

目标分布是用核密度估计算出来的，当n取50时，是一个50\*50的二维矩阵，代表一个二维分布。风功率曲线是风速和风机发电量的二维分布，w代表风速，p代表风机发电量 。模型输出的风功率曲线是f(wreal, pmodel)，真实风功率曲线是f(wreal, preal)，在使用核密度估计时需要判断待估计变量x的上下界，生成一系列区间[x1, x2, ..., xn]，n为区间数，因为研究中需要让模型去拟合一个分布，一个很直接的方法是直接取模型第一次输出值的最大最小值作为核密度估计的上界和下界，这样的结果是在训练完成时模型无法对真实数据的上下界做出限制，只能在形态上让两个分布一致，如下图所示，

如果采用同一上下界对模型和真实的风功率曲线进行对比，会发现模型缺失了一部分的统计信息。因此在训练时，不能动态地调整上下界而应该始终使用真实的上下界。这样引入了一个新的问题，在神经网络的的训练过程中，模型的输出刚开始会集中分布在一个很窄的区间，如果采用固定的上下界，模型的输出在很大的范围可能不会参与统计，在计算分布时，没有数据的地方我们不会将概率设置为0而是一个很小的值，但任然会出现统计结果无法求梯度的情况，从而导致模型无法更新甚至训练崩溃，



为了解决这个问题，又引入了js\_ratio这个概念，就是为了解决神经网络在训练初期无法得到有效统计信息的情况，如果模型输出结果无法统计，则将js\_ratio置0，这个时候由mse loss来训练神经网络，知道模型可以输出一个可以统计的值，在将js\_ratio置1，使js loss参与到网络训练。这只是一个方法，其他的方法包括使用一个渐进值，在训练初期将js\_ratio置0，随着迭代次数将js\_ratio逐渐加大。

新的问题还有js loss波动性往往较大，在模型收敛的时候，js loss的波动幅度仍然很大，导致测试结果相较于mse训练出来的模型波动更大。这个需要进一步讨论，在本研究中只讨论js loss的优化效果。

源奇您好，我把背景和方法部分的初稿写得差不多了了，正准备写实验部分，还有几个问题想向您请教下：1、发动机转速这个变量是否用到，如果没用的话为什么没用？

2、为什么用公式（4）这种方法选择带宽？公式中n代表什么，怎么选的？

3、js\_ratio怎么选取的，在代码里面有没有体现？

4、为什么要加N(x,x)的噪音？训练噪音的加入在代码里面是怎么体现的？

5、图10和11表格里的值是什么含义？

6、最后一个表格表1是什么实验设置？

7、代码里面怎么用到的天气预报数据，相关结果是否有记录？

1. 没有用到发电机转速这个变量，因为发电机转速跟发电量功率高度相关，近乎线性关系。
2. 公式（4）中n为直方图分箱数量，当n取50时表示用50个区间来统计变量。本次研究中只选择一个n，具体数值可以根据统计需要来调整。
3. 引入js ratio的目的是调节训练过程中js loss的比例，由于一开始模型输出功率的统计特征不太明显，无法计算出风功率曲线值，当出现这种情况时，js loss往往会出现没有梯度的情况（如模型初始输出非常均匀，在统计上会集中在一个很小的范围，无法计算模型输出风功率曲线和目标风功率曲线的js 距离）。出现这种情况可以根据模型输出的方差来判断模型的输出是否可以统计。如果无法统计则将js ratio设为0，即不用js loss，而使用mse来对网络进行训练，直到模型可以输出一个可以统计的功率，js ratio设为1，再将js loss参与到训练当中。代码中可以查找js\_ratio这个变量的相关代码段。这个部分需要细致讨论。
4. 因为训练和测试如果都是实测数据，实验则有些单薄，无法对应真实功率预测下未来风速不可知的环境，因此对风速添加一些噪音模拟预测风速。N（x，x）噪音的加入是为了体现一种有偏的噪音，即会使添加噪音后的风速在统计量上发生偏移，这样能体现出js loss的优势，因为只要是统计量的偏移，都会对js距离产生较大影响，而mse无法在统计上对数据进行拟合。
5. 图10是N（0， x）噪音的实验结果，图11是N（x， x）噪音的实验结果，每一个大的行如N（0， 0.1）表示在训练中添加了这种噪音的实验结果，每一个大的列如0.1表示在测试中添加了N（0，0.1）的噪音，小行中10，20，30，40，50表示跑了多少个epoch
6. 表格1的实验设置是将风速替换成真实的预报风速，即训练和测试中使用的都是预报风速。在模型训练过程中，训练数据中输入的风速是3天历史风速加上一天未来的风速，这里直接将未来的风速用真实预报风速替换。
7. 代码中time\_ahead()函数中可以对数据进行切换，因为上次问题中的问题3里的wp就是预报风速，可以在拼接训练数据的过程中将其替换。

源奇您好，我把代码在我电脑上跑通了，目前在研究代码，有几个问题想向您请教下：1、JSFD和江苏十分钟数据的关系？训练的时候是不是就用不到JSFD和JSFD\_MERGED文件了？

2、summary.py里面pc\_record和load\_record在哪？

3、train\_dayahead.py里面的wp在哪里用到了，和那个”平均风速“的关系是？

4、AutomaticWeightedLoss有没有用到？

5、template的图1和图2代表什么（或者说明下画这些图的代码在哪，我可以自己去研究下代码）？

1. JSFD 是预报数据，时间上可以和江苏十分钟数据对齐，作为真实数据的对比，训练中可以作为一个特征去使用，可以在代码中构建数据集时添加进去。
2. summary.py缺的函数已经补齐
3. wp是JSFD的风速预报数据
4. AutomaticWeightedLoss没有用到可以删除
5. 图一是summary.py中生成的pc\_summary.pdf截图意思是训练过程中模型输出的风功率曲线，图二是直接在终端画的具体意思是estimate是核密度估计算出来的风功率曲线，true是基于统计方法算出来额风功率曲线，residuals是ture-estimate。

