HPCA-AI 博士会议2019-10-08

会议记录：王昕

今天和薛老师、李老师一起研讨学习了目前气候模式参数优化工作的状况。

**（一）气候模式参数优化背景**

首先是整个参数优化研究的发展趋势，由单目标优化向多目标最优化发展。

在整个参数优化研究中，需要获取海量数据样本。

因而，计算成本问题是一个首先需要解决的问题。

· 成本最高的是参数采样

· 对参数进行筛选，以降低总计算成本

· 参数变化与预测误差之间的关系呈现出高度非线性

参数组的设置链，每次参数变化是基于前一次参数的，有数据依赖性。只要这条链中有一个参数出问题，整个链条的运行就 crash 了，产生的数据也就没用了，需要重新进行模拟。

为了解决参数优化过程复杂、计算量大，数据采样成本高的问题。目前采用以下方式:

1.在工程上，可以对采样链的进行重新规划，把容易崩溃的采样环节放于采样链的前端，并挖掘其中的可并行性;

2.通过构建 CAPT 优化采样过程;

3.使用精简快速的代理模式;

4.动态参数筛选，筛选出对结果影响较大的敏感参数（组合）。

**（二）CAPT - 快物理过程偏差的快速诊断方法**

根本原理:

快过程产生了一个比较大的偏差；慢过程产生了一系列的、累积起来的偏差。通过寻找到快过程，可以确定产生大偏差的原因，进而缩小参数优化的搜索空间。

CAPT 特点:

1.快慢过程产生偏差的原因是模式本身的问题，与参数关系不大。

2.使用快过程，可以加速（优化）采样的过程。

3.CAPT 可以解决短期产生明显偏差的参数，但无法反映出通过长期运行，积累产生偏差的参数。

总结: CAPT 同步方法优化采样过程来完善整体的参数优化过程。

**（三）代理模式**

总结:

    1. 用已有的数据，或自己额外做模拟所获取的数据集构建代理模式；

    2. 代理模式类似于自然降维；

    3. 没法用先验的方式来判断该代理模式是否是过拟合还是欠拟合；

    4. 没有一个评价代理模式总体性能的指标；

    5. 把代理模式和采样结合在一起，可以使得原模式的采样更稀疏，降低计算量。

代理模式发展历程:

      1.仅仅通过RBF 构建代理模式；

—> 2.多目标代理模式（不同的目标选择不同的构建代理模式的方法，比如降雨使用 RBF）；

—> 3.通过 DBN（进行降噪，在不同目标之间，再）结合 RBF 构建代理模式；

—> 4.完全使用神经网络去构建代理模式；

构建代理模式是目的是优化采样过程。

**（四）动态参数筛选**

参数的敏感度不是孤立的，是与其他的参数相关联的。目前通过图来表示参数之间的敏感关系。

参数之间的敏感度关联无法直接体现，必须通过模式运行后所产生的误差才能得以体现。

「稀疏问题」是如何与「参数之间的关系」所联系到一起的？

只要是强相关领域，就可以利用 Sparse coding，但参数不是纯粹的数据，其相关性也无法直接从参数本身得到反应。

如何在动态参数筛选中引入稀疏处理很重要

总结: 动态参数筛选来优化参数的数量，减少需要优化参数空间。

**（五）模式预测的动态变化**

在气候模式中，随着时间的变化，输入数据参数都会改变。现有的方式大多是静态处理这些变化。

如何进行动态处理，将是接下来的重要研究方向。