调研报告13

PM2.5 issue

(2020.2.13)

**一、problem**

假设存在这样一个问题：

给定中国400多个城市的当地PM2.5数据，

1. 这些城市的pm2.5数据在空间上肯定是有关联的，因此构成了一个多任务学习，多个任务之间肯定有共同的特点。

2. 但是受地形，大气环流影响，这些地方的pm2.5时间序列肯定具有当地的变化特性，那么每个地方之间的任务模型受本地因素影响也比较大。

3. 当地的task-specific不止受当地影响（2中所述），当地的模型参数应该还与附近最近的几个地点之间存在关联关系。

因此还存在一个小范围，局部的影响关系。

综上所示，整个模型受三个层次的影响，

1）最高层次的整体的multitask，400多个城市的共同特征。

2）当地小范围的multitask 影响关系，大概3-5个task

3）本地的污染排放等，纯task-specific的影响。

**二、idea**

**（一）比较符合上述问题的模型：利用树的结构（Tree-structure）**

**1. 论文**

Han, L., & Zhang, Y. (2015). **Learning Tree Structure in Multi-Task Learning** Categories and Subject Descriptors. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 397–406.

**2. 基本思想**

该论文利用树的结构来描述任务之间的层次关系。

该论文致力于研究：

·如何给出（define）多任务树结构（task tree）？

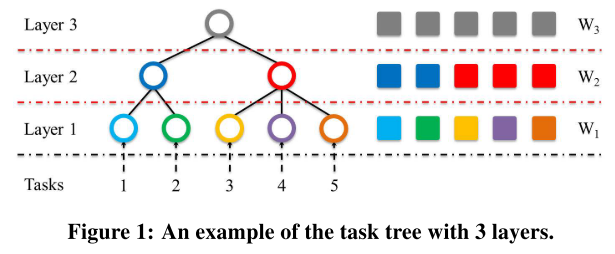
提出TAT(TAsk Tree)method to learn the underlying task tree and the model parameters simultaneously

·如何解决模型中的非凸？

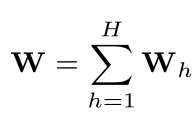
利用general iterative shrinkage and thresholding (GIST) method，其中利用ADMM method, 将GIST method中的问题分解成三个独立的子问题

树的结构的定义：

每一个叶子节点是一个task，当两个任务之间有关系，则他们在同一个子树上，共享一个子树根节点。



W是参数矩阵，**在每一层（layer）中分解该参数矩阵**。树共有H层，每一层分解后的矩阵为Wh.

即，

其中每一层（layer）有，wh,i代表第h层的第i个任务的参数。

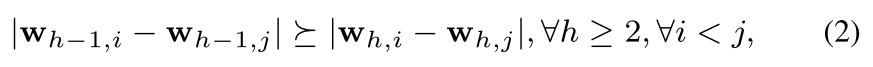
**定义**：对于，满足下面两个条件成为一个任务树，

1. 不同任务对应第h层的参数都存储在子矩阵Wh中.
2. 对于任意一对任务，例如第i和第j个任务，如果第h层对应的参数满足

wh,i = wh,j，则wh’,i = wh’,j ,对任意h’≥h。（就是说，在h层两个任务的参数相等，那么层数再往上走，靠向总树根节点，那么这个两个任务的参数一定是相等的。）

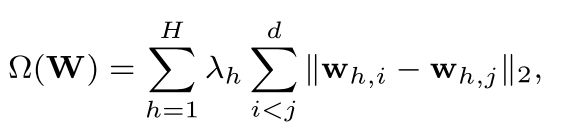
两点说明：

1. 为了满足定义第②条，则规定



这样的序列限制条件，就能保证满足②。

2）为了保证第i和第j个任务参数能够融合到一起，wh,i = wh,j，定义正则化项

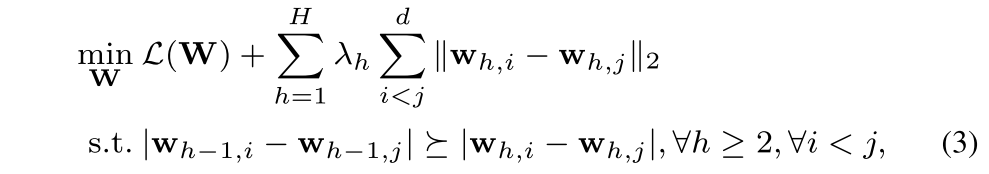


Note:

λ1, …, λH ：to specify the importance of different layers.

L2 范数的Ω(W) 趋向于使wh,i = wh,j.

**提出的TAT模型：**



Note：λh控制了第h层任务的相似的强度，λh越大会导致矩阵中任务参数wh,i和 wh,j越相近。本论文中定义使用λh =λh−1，h ≥ 2，常数> 1.

本论文定义的loss function为



论文接下来利用GIST解决由于正则项非光滑和约束条件非凸产生的问题。

**3.接下来的深追**

·一些层次结构模型（hierarchical models）：论文的参考文献[15, 9, 21, 28, 12]。

**4.问题**

1）关于tree-structure multitask learning的论文比较少，而且引用量不是很多，这个方法是否靠谱，是否应该继续深追下去？

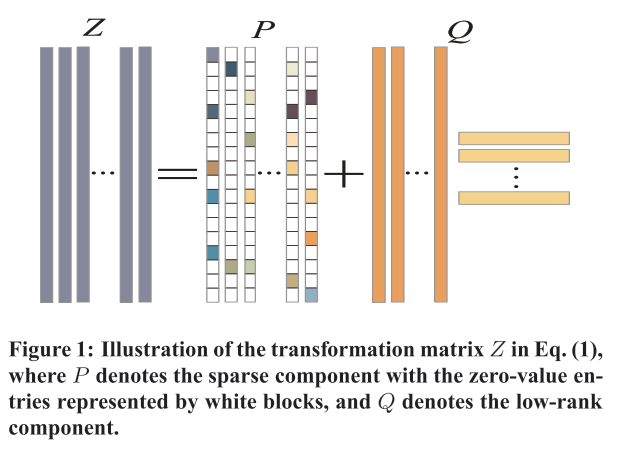
2）我的想法，对于PM2.5的问题，将整个任务树分三层，实现PM2.5 problem的三个层次，将400多个城市视为400个tasks（叶子节点），利用task group或者task clustering的方法，将地理上相近的3~5个城市共用一个子树，最后总树的根节点代表整个400个城市的共同特征。

3）多任务学习，现在看的多是结构式的方法，类似多任务和深度学习结合的比如硬参数共享、软参数共享和神经网络这类的论文是否要看？能否成为我们的研究方法？

**（二）稍微有点欠缺的模型：**

再有看了一些论文，大多都是两层的模型，比如下面这篇。

Chen, J., Ji, L., & Ye, J. (2010). **Learning incoherent sparse and low-rank patterns from multiple tasks**. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, May 2014, 1179–1187.



Z=P+Q

矩阵P是稀疏的Sparse：代表任务之间的差异

矩阵Q是Low-rank：代表任务之间的底层潜在性