总结报告14

Distributed Learning of Predictive Structures

from Multiple Tasks over Networks

（2020.2.14——2020.2.15）

一、论文总结

Hua, J., Li, C., & Shen, H. L. (2017). **Distributed Learning of Predictive Structures from Multiple Tasks over Networks**. IEEE Transactions on Industrial Electronics, 64(5), 4246–4256.

（一）idea

本论文认为multi tasks之间有一个low-dimensional predictive structure。学习这些tasks通过graph的network。各个节点之间通过图连在一起，没有中心节点(分布式与联邦的不同，一般联邦都有master聚合参数)。

1. 论文脉络

1．Network model

Graph。每个node的liner predictor



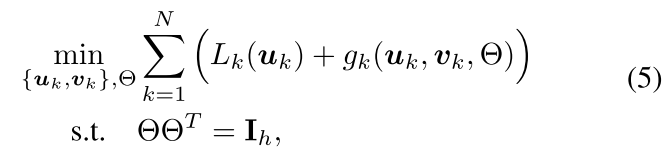
2．Formulation of Distributed Multitask Learning



将（1）拆成两部分，是权重向量，要从每个节点计算。即是作者提出的common structure，但是有一个正交的限制条件。

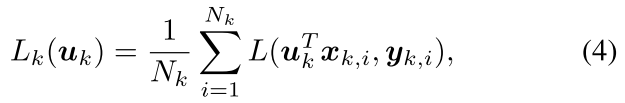


接下来得到了整个模型，

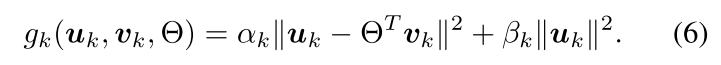


Note：

Lk：loss function



gk：regularization function



where ：

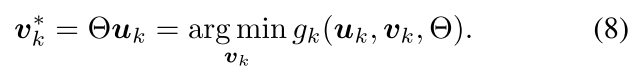


本论文中假设了所有α都相等，β也都相等。

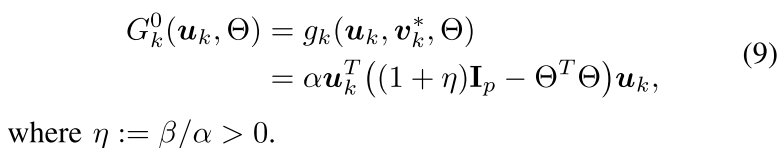
3. Convex Relation

由于（5）正交的限制条件，是非凸的。

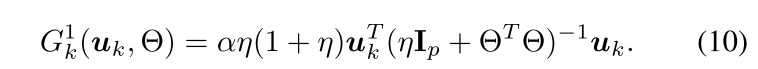
（5）中的的最优解为



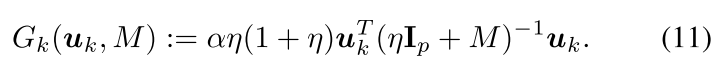
消除，重写正则项为，



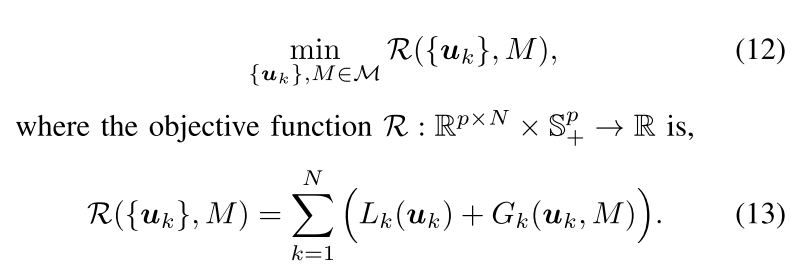
由(1 + η)Ip − ΘTΘ = η(1 +η)(ηIp + ΘTΘ)−1进而正则项变为，



由推导，放松ΘTΘ为M，最后正则项变为，

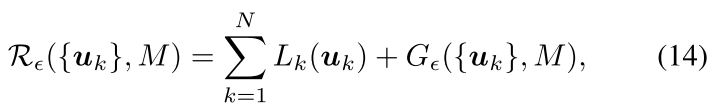


整个目标函数变为，

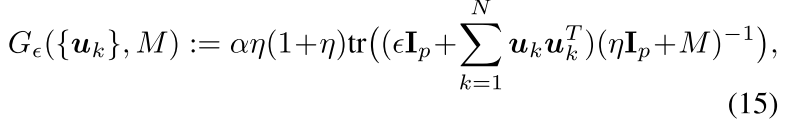


其中，见（11）

考虑扰动，加入了一个系数，最后的目标函数为，（w.r.t 和M）



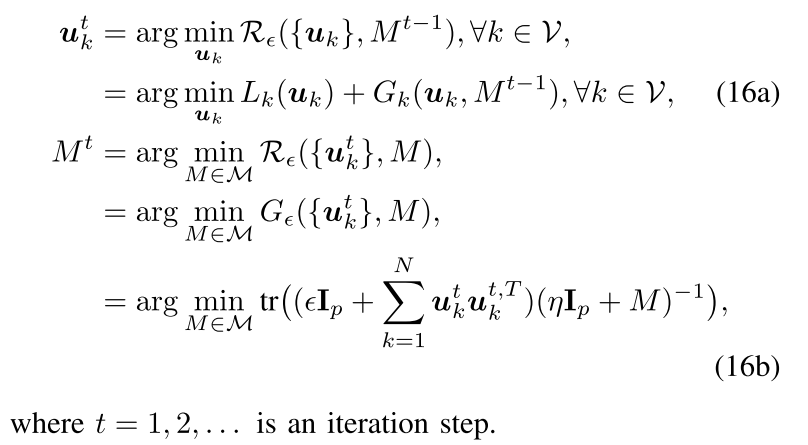
其中，



4. Distributed Multitask Learning

4.1 Centralized Multitask Learning

中心式多任务学习利用的是BCD方法：



4.2 Distributed Multitask Learning

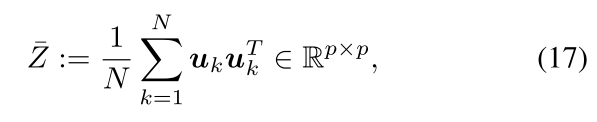
分布式体现在这：

是每一个node在本地算出来的，但是（16b）展示了M的计算需要每个节点的，势必要集中。

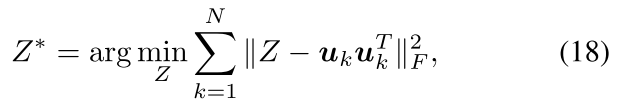
这里作者通过实现了分布式，即将M的计算分配给了每个node。

具体分布式实现推导：

可以看到（16b）依赖一个中间量，

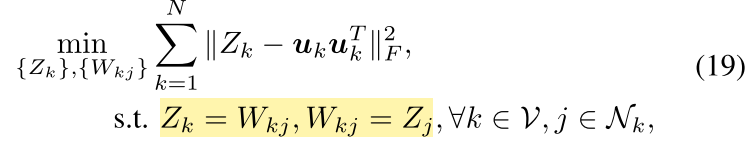


（17）等价于（18），



给每个节点定义出Zk，，添加一致性约束来强制Z在相邻节点之间相等。

即，



Note:是辅助变量。

可以证明，最后每个Zk 都相等，等于Z。

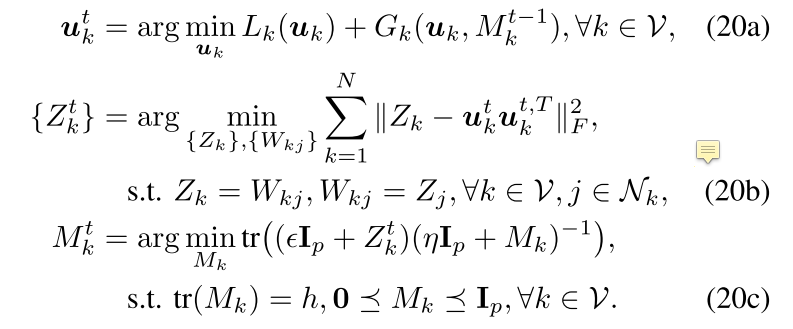
点评：

原来是关于和M的目标函数，作者提出了中心式多任务算法，再将中心式多任务算法逐步改造成分布式算法。这里面最主要的就是因为M的计算对分布式造成了障碍。于是

分布式体现在这：



分布式多任务学习算法：



Note：（20a）（20c）只是节点自己完成，（20b）需要和邻居合作。

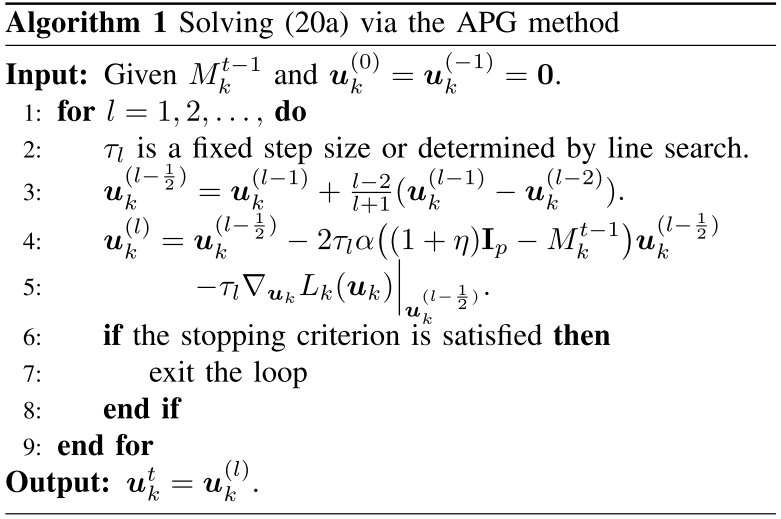
这个算法有两层循环。外层的循环（iteration *t*）利用BCD算法（中心式算法即BCD方法，分布式外壳是套过来的），后面介绍到内层循环（iteration ）解决（20b），利用ADMM。

5. 下面是就分布式多任务学习算法的每个step进行计算：

5.1 （20a）的解决

利用APG（Accelerated gradient descent）method。

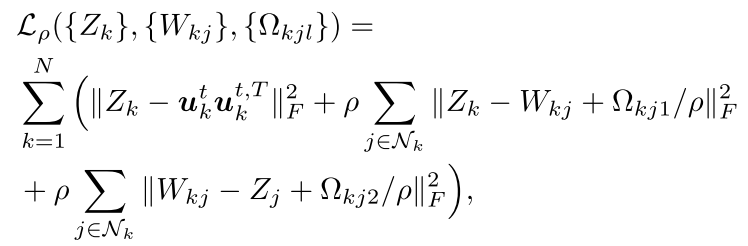
Mk利用上一次的最优值。



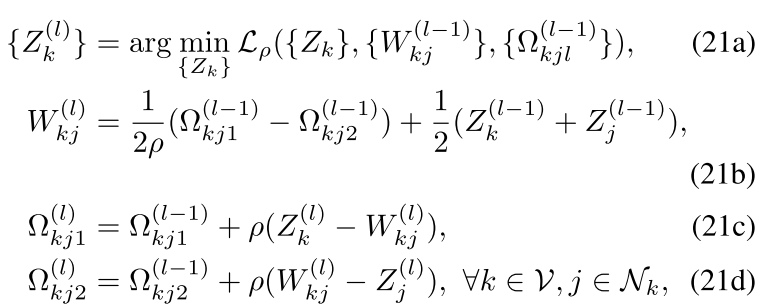
5.2 （20b）的解决：Zk的分布式计算

利用ADMM。（增广拉格朗日、对偶上升）

（20b）的Lagrangian function：

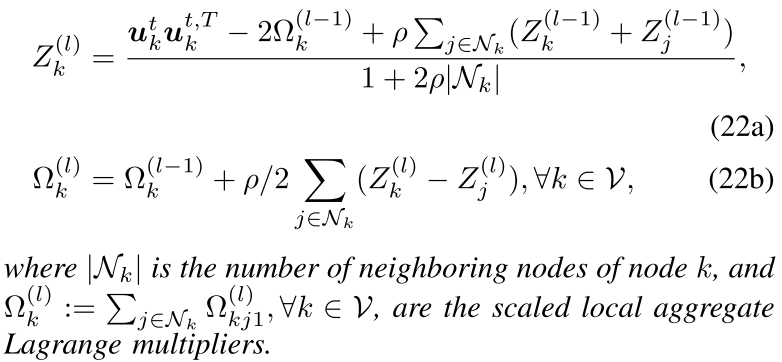


Note:Ωkj1和Ωkj2:Lagrange multiplier对应约束Zk = Wkj 和Wkj = Zj



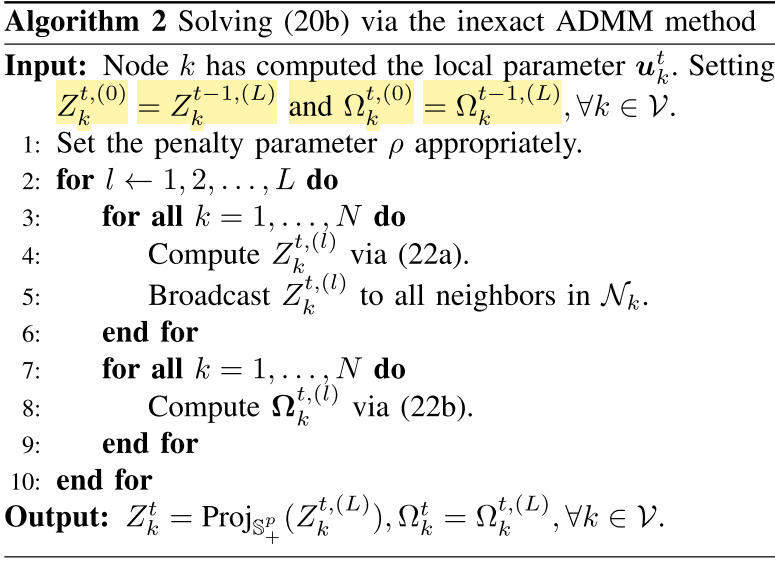
Note：其中，Wkj是辅助量

（21a—21d）可以进一步简化为：



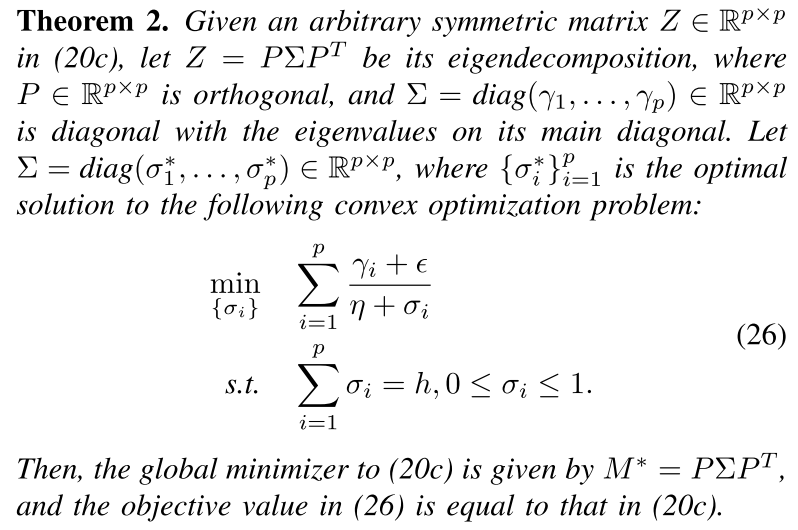
（20b）需要和邻居nodes合作，为了减少交流成本，采用一个非精确方法代替。

Summary：解决（20b）。注：标黄的表示不精确方法代替

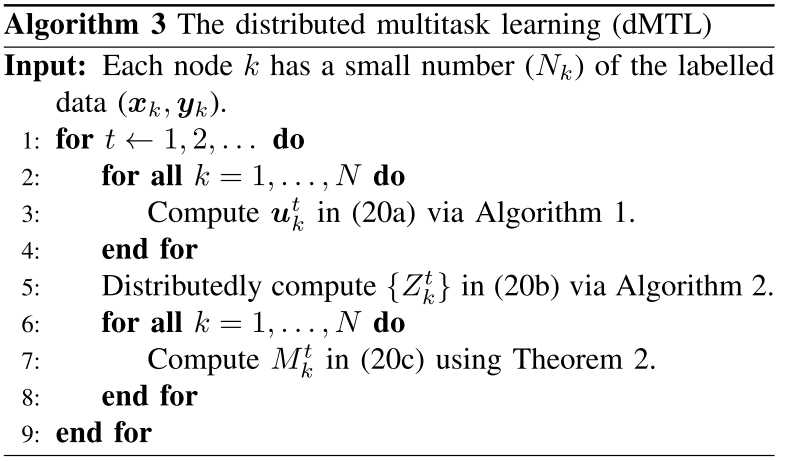


5.3 Local Computation of Mk in（20c）：解决（20c）Mk的算法

通过特征值优化问题得到结果Mk。矩阵分解（证明见这篇论文的参考文献[21]）



（三）总结算法dMTL



二、其他论文

Lozano, A. C., & Świrszcz, G. (2012). **Multi-level Lasso for sparse multi-task regression**. Proceedings of the 29th International Conference on Machine Learning, ICML 2012, 1, 361–368.

两层，让y=β\*x的β分解为\*γ，而不是以往的β分解为相加形式。表示全局，γ是task-specific。两个L1-norm即multi-lever lasso

Han, L., & Zhang, Y. (2015). **Learning multi-level task groups in multi-task learning**. Proceedings of the National Conference on Artificial Intelligence, 4, 2638–2644.

和Learning Tree Structure in Multi-Task Learning这篇论文完全一样，作者都一样

Fiedler, M. (n.d.). **Privacy Preserving QoE Modeling using Collaborative Learning**. 0–5.

只是基本提出了协作学习的方法：

协作学习的两种：RRL，节点之间相互share模型，以循环的方式ring all reduce；FL，有一个中心master，把得到的weight给到master node，master更新后再传播给各个节点

Lalitha, A., Cihan, O., Tara, K., & Farinaz, J. (n.d.). **Peer-to-Peer Federated Learning on Graphs**.

不需要master，直接在nodes之间交换权重

Fficiency, C. O. E., Mcmahan, H. B., Yu, F. X., Suresh, A. T., Bacon, D., & Richt, P. (2017). **Federated Learning: Strategies For Improving Communication Efficiency**. 1–10.

联邦学习的基本概念