总结报告18

Distributed Multi-Task Learning with Shared Representation

（2020.2.27）

一、contents

Wang, J., Kolar, M., & Srebro, N. (2016). **Distributed Multi-Task Learning with Shared Representation.**

Takahashi, N., Yamada, I., Member, S., & Sayed, A. H. (2010). **Diffusion Least-Mean Squares With Adaptive Combiners : Formulation and Performance Analysis**. 58(9), 4795–4810.

Jie Chen, Member, IEEE, Cédric Richard, SeniorMember, IEEE, and Ali H. Sayed, Fellow, I. (2016). **Multitask Diffusion Adaptation Over Networks**. IEEE Transactions on Signal Processing, 64(11), 2835–2850.

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

二、论文1

1. 题目

Wang, J., Kolar, M., & Srebro, N. (2016). **Distributed Multi-Task Learning with Shared Representation.**

2. 内容

2.1 idea

所有machines共享一个shared low-dimensional subspaces（the predictor matrix是低秩的）。

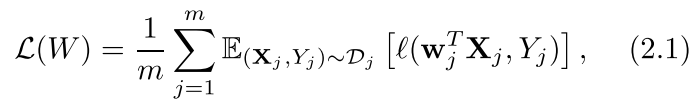
**最大看点**在于：利用了贪婪算法来得到低维的subspace，并且subspace的这些向量都是互相正交的。

2.2 符号说明

一共有m个machines（tasks），每一个machine有n个examples

2.3 目标函数

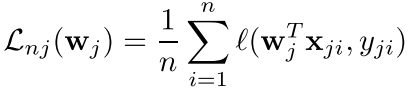
Global empirical loss：



Distributed empirical loss：



其中，



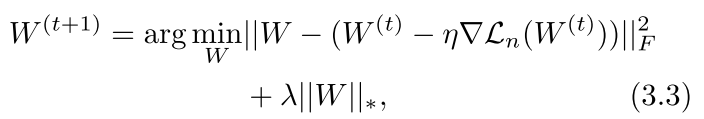
把低秩的条件加进去作为正则项，

包括正则化的objective function：

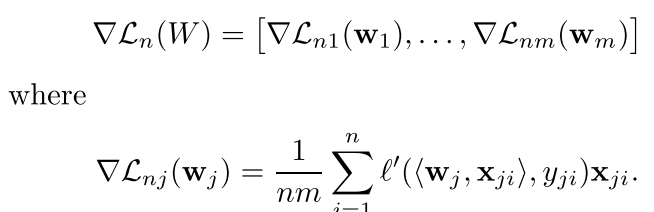


2.4 问题（3.2）解决方法1——distributed proximal gradient

利用近端梯度下降（distributed proximal gradient），



其中，



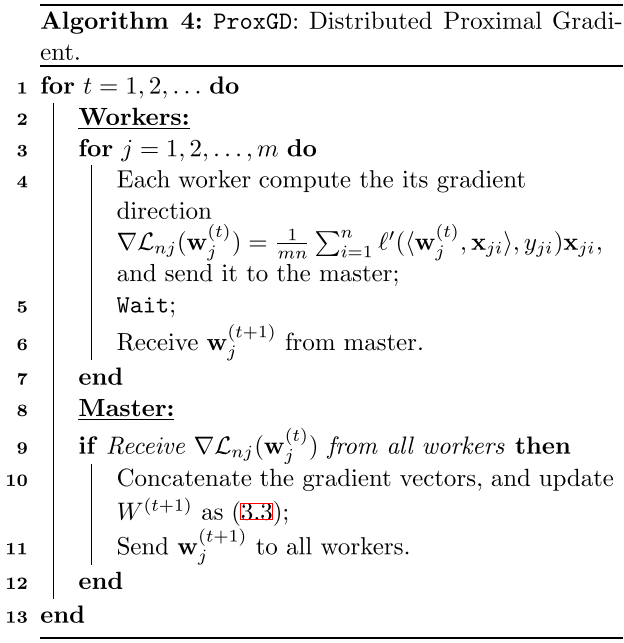
对 进行SVD分解，



那么，

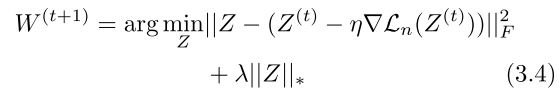


其中，



2.5 问题（3.2）解决方法2——distributed accelerated gradient

引入Z，

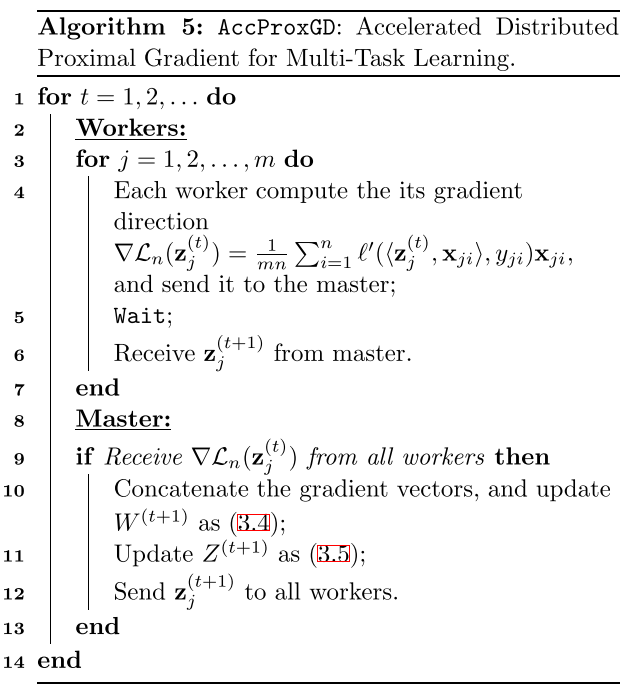




Accelerated与没有Accelerated的比较：

——收敛：从2.4的到2.5的

——交流的rounds：从2.4的到2.5的



2.6 Greedy Representation Learning贪婪

* 是低秩的，rank=r.

把写成r个秩为1的向量和，即u。

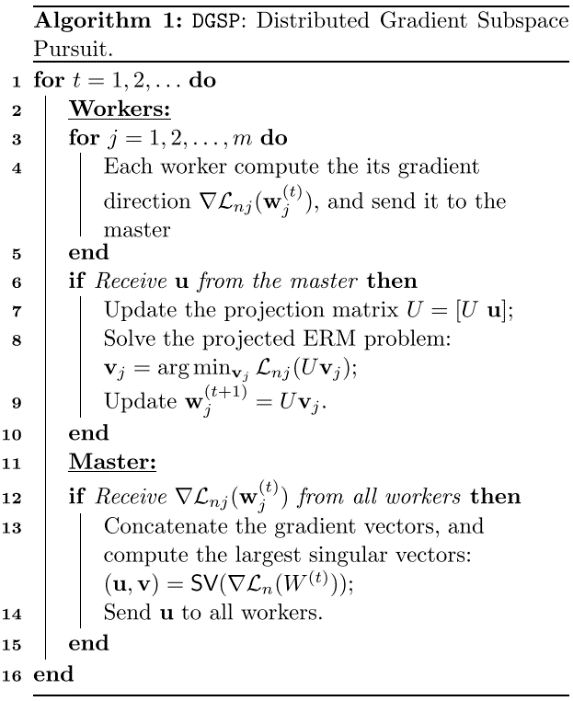


Note：把W表示乘U和V的向量乘积。

U：projection matrix

V：the coefficient vectors



Note:

当master得到所有node发送的时，组合成一个向量，它的最大的奇异向量为，把u加入到U中，每一次迭代得到一个u，即贪婪。

三、论文2

1. 题目

Takahashi, N., Yamada, I., Member, S., & Sayed, A. H. (2010). **Diffusion Least-Mean Squares With Adaptive Combiners : Formulation and Performance Analysis**. 58(9), 4795–4810.

2. 内容

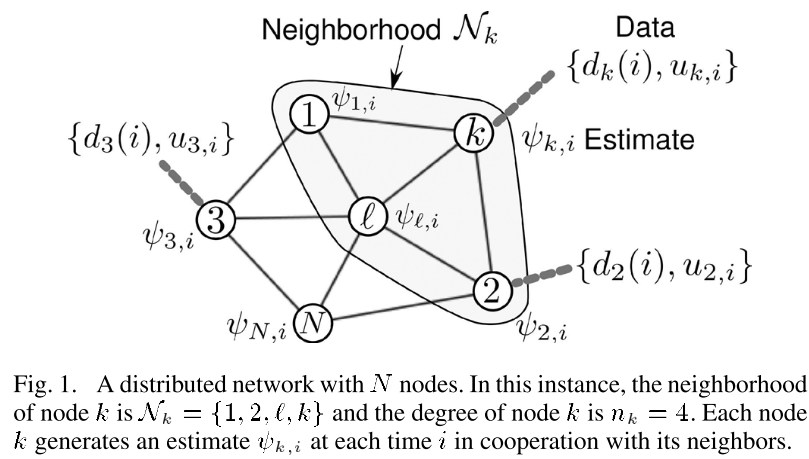
2.1 重点

这篇论文主要看点是Adaptive Combiners：因为邻居之间要交流，所以，Adaptive Combiners就是讲，如何将邻居们交流来的权重向量聚合起来。并介绍了两种传播策略（先传播谁，后传播谁）

目标函数和求解方法都不太合适，就不整理了。

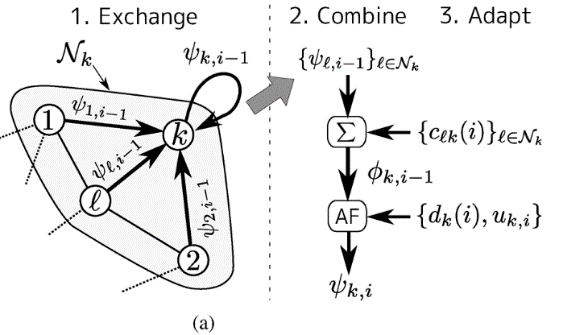
2.2 符号说明

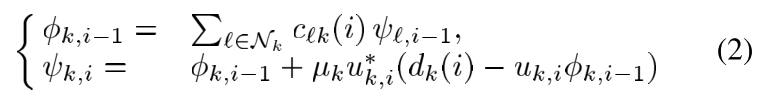




2.3 这里首先两种传播策略diffusion strategies

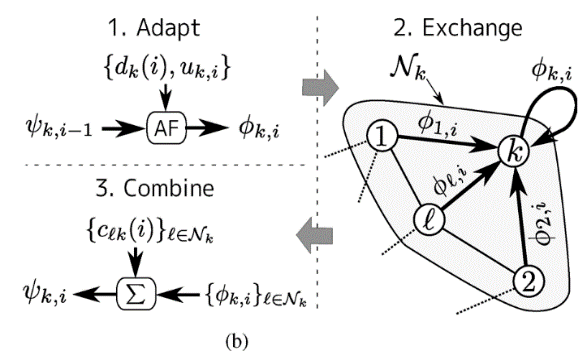
（一）CTA：Combine-then-Adapt

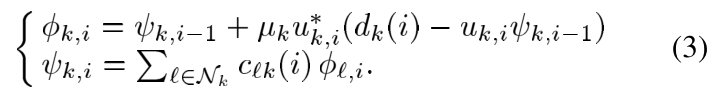


Note:表示共轭

（二）ATC：Adapt-then-Combine

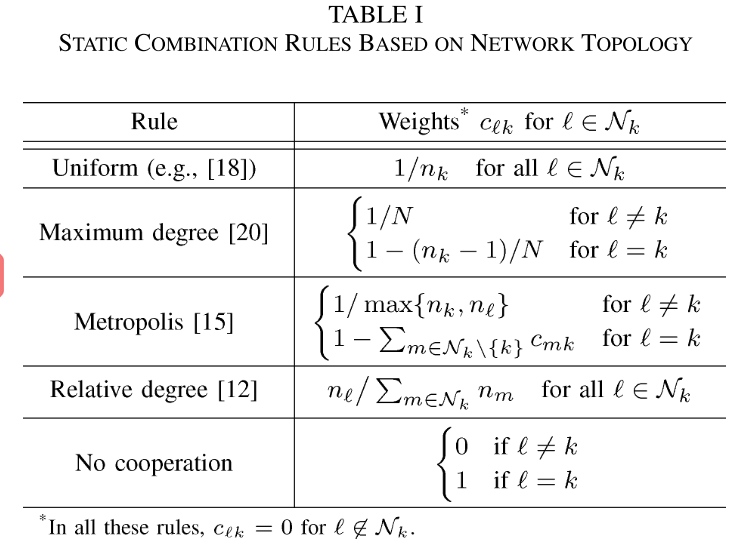






2.4 Adaptive Combiners：

的取值方法有如下几种，以便聚合邻居的参数w。

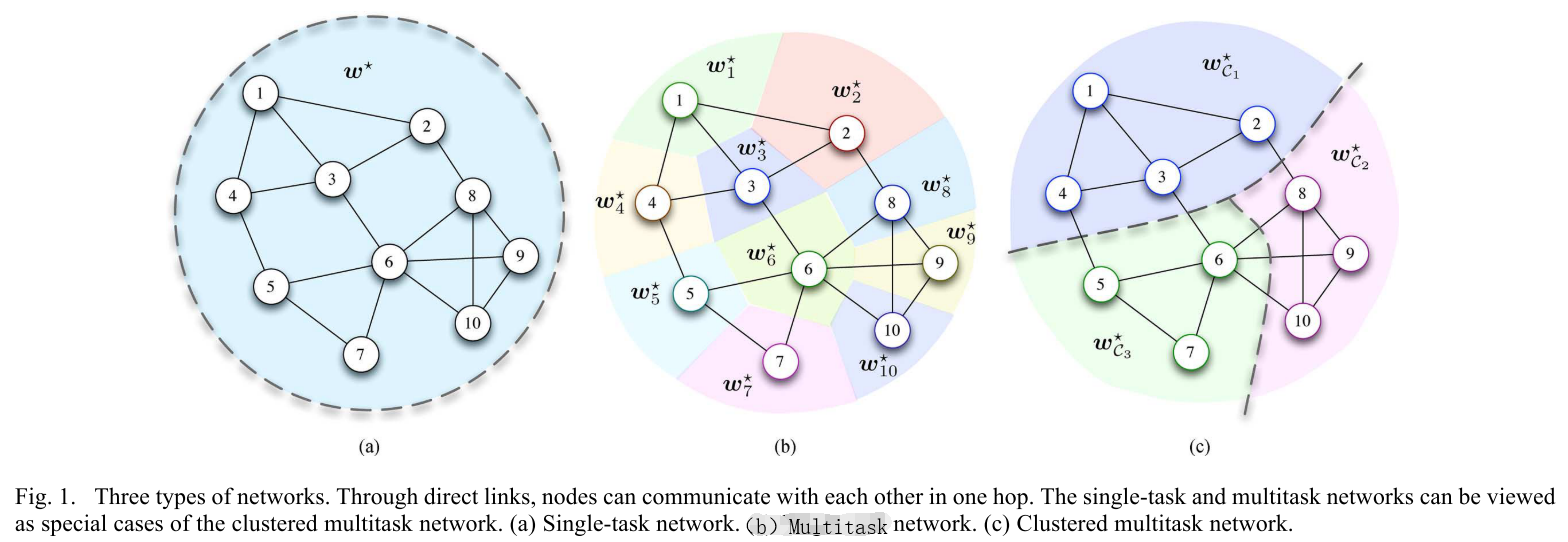


四、论文3

1. 题目

Jie Chen, Member, IEEE, Cédric Richard, SeniorMember, IEEE, and Ali H. Sayed, Fellow, I. (2016). **Multitask Diffusion Adaptation Over Networks**. IEEE Transactions on Signal Processing, 64(11), 2835–2850.

2. 内容



作者想到达到的状态是：以上图（c）为例，想让同一个cluster内的直接近邻（直接有一条边相连）之间有强的信息交流。不同的cluster之间也有比较弱的交流，而这种交流仅当两个点位于不同cluster但两点间有一条边之间相连才会发生。作者在文中用了极其繁琐的符号表示，但是实际到最后，又进行了近似等等，变得可解。

另外，作者引入了博弈论中的纳什均衡（Nash equilibria）（所谓纳什均衡，指的是参与人的这样一种策略组合，在该策略组合上，任何参与人单独改变策略都不会得到好处。）实际就是alternating method，但是引用了纳什均衡的说法显得高级些？

五、论文4

1. 题目

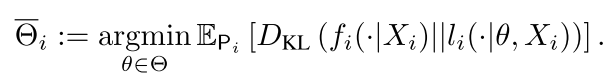
Lalitha, A., Cihan, O., Tara, K., & Farinaz, J. (n.d.). **Peer-to-Peer Federated Learning on Graphs**.

2． 内容

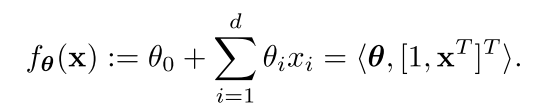
2.1 idea

对全局模型的分布式学习是通过两个步骤来实现的。首先，每个节点根据基于节点本地数据的**贝叶斯推理**步骤(后验更新)更新其本地信念（有点概率的意思，local belief ρ）。然后，聚合来自单跳邻居（one-hop）的信息。得到（global belief b）

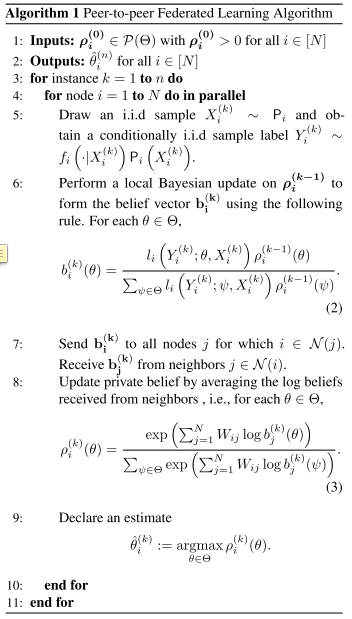
2.2 目标函数



Note:

：KL散度，度量的是样本的概率分布与真正总体的概率分布的差别。

2.3 算法

Note：

跟概率结合起来，贝叶斯推断。

Local belief ρ

Global belief b

Line 7：由于是完全分布式的，所以没有中心节点。