**Distributed Systems**

**（2020.12.28）**

1. 可参考的报告

总结报告11、12、14、15、16、17、18、22、24、28

1. 可参看的论文

|  |  |  |  |
| --- | --- | --- | --- |
| 报告 | 论文题目 | 算法缩写 | 一句话总结(related works) |
| 11 | Liu, S., Pan, S. J., & Ho, Q. (n.d.). **Distributed Multi-Task Relationship Learning.** 1–31. | DMTRL | 中心分布式。更新权重(W, local)，更新任务关系(task relationshipΣ, center) |
| 12 | Smith, V., Chiang, C., Sanjabi, M., & Talwalkar, A. (2017). **Federated Multi-Task Learning.** Nips. | MOCHA | 中心分布式。通过一个参数，在一个iteration中，可能有的节点就不发送参数给center |
| 14 | 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. | dMTL | 完全分布式。从单一节点算法逐步改造成分布式算法。有两个参数u，M。u很容易改造成分布式，但M不易，故通过逐步改造实现M的分布式计算。 |
| 15 | Wang, W., Wang, J., Kolar, M., & Srebro, N. (2018). **Distributed Stochastic Multi-Task Learning with Graph Regularization**. | Distributed algorithm for ERM | 完全分布式。引入Laplacian 图矩阵，先算出global problem的闭式解后分配到nodes，研究重点在于通信方式、效率、成本。 |
| Li, C., Huang, S., Liu, Y., & Zhang, Z. (2018). **Distributed jointly sparse multitask learning over networks**. IEEE Transactions on Cybernetics, 48(1), 151–164. | AC-dJSMT (adaptive *clk*distributed jointly sparse multitask) | 完全分布式。研究邻居的通信方式，通过参数*clk*intertask combiners，决定是否邻居间是否交换参数。 |
| Verma, V. K., & B, P. R. (2017). **Distributed Multi-task Learning for Sensor Network**. 1, 792–808. | DMTL | 完全分布式。中规中矩的在neighbors间传递参数parameter |
| 16 | 同15（修改增添） | — | 可不看报告15，直接看报告16 |
| 17 | Zhou, Q., Chen, Y., & Pan, S. J. (2019). **Communication-efficient distributed multi-task learning with matrix sparsity regularization**. Machine Learning. | Efficient Distributed Optimization | 中心分布式。不按data分，而是按feature分配给nodes。为减少通信成本，提出了dynamic feature screening，识别inactive feature，不参与通信 |
| 18 | Wang, J., Kolar, M., & Srebro, N. (2016). **Distributed Multi-Task Learning with Shared Representation.** | ProxGD  AccProxGD  DGSP贪婪 | 中心分布式。核范数在center更新，ERM分配给节点，返回梯度给center，不同算法以提高node的计算速度 |
| Lalitha, A., Cihan, O., Tara, K., & Farinaz, J. (n.d.). **Peer-to-Peer Federated Learning on Graphs**. | Peer-to-Peer Federated Learning Algorithm | 完全分布式。跟概率结合起来，贝叶斯推断。 |
| 22 | Li, W., Liu, Y., Tian, Z., & Ling, Q. (2020). **Communication-Censored Linearized ADMM for Decentralized Consensus Optimization**. IEEE Transactions on Signal and Information Processing over Networks, 6(1), 18–34. | COLA | 完全分布式。有一个阈值，只有变量的变化大于阈值，才开始进行交流，弥补了线性化步骤造成的通信效率低下的问题 |
| Richtárik, P., & Takáč, M. (2016). **Distributed coordinate descent method for learning with big data**. Journal of Machine Learning Research, 17. | Hydra: HYbriD coordinAte descent | 完全分布式。坐标下降，每次迭代只有几维更新，并非全部更新。 |
| Duchi, J. C., Agarwal, A., & Wainwright, M. J. (2012). **Dual Averaging for Distributed Optimization: Convergence Analysis and Network Scaling**. IEEE TRANSACTIONS ON AUTOMATIC CONTROL, 57(3), 1564–1565. | dual averaging algorithm | 完全分布式。无正则项。每一个node都计算它的次梯度，从它的邻居节点得到z(t)后加权平均，之后再计算它的下一次更新x(t+1)，利用一个（由和α(t)定义的）投影。 |
| 24 | Jaggi, M., Smith, V., Takáč, M., Terhorst, J., Krishnan, S., Hofmann, T., & Jordan, M. I. (2014). **Communication-efficient distributed dual coordinate ascent**. Advances in Neural Information Processing Systems, 4(January), 3068–3076. | CoCoA | 完全分布式。Nodes间只需要通信一个单一的向量。 |
| Smith, V., Takáč, M., & Jordan, M. I. (2018). **CoCoA : A General Framework for Communication-Efficient Distributed Optimization.** 18, 1–49. | Generalized  CoCoA | 完全分布式。更泛化，能解决non-strongly-convex regularizers and non-smooth loss functions |
| Ma, C., Smith, V., Jaggi, M., Jordan, M. I., Richtárik, P., & Takáč, M. (2015). **Adding vs. averaging in distributed primal-dual optimization**. 32nd International Conference on Machine Learning, ICML 2015, 3, 1973–1982. | CoCoA+ | 完全分布式。  修改了subproblem，更有效把子问题结合；  A aggregation parameter γ收敛更快。 |
| Smith, V., Forte, S., Jordan, M. I., & Jaggi, M. (2015). **L1-Regularized Distributed Optimization: A Communication-Efficient Primal-Dual Framework**. 1–23. | PROX  CoCoA+ | 完全分布式。改进在减少通信成本。 |
| 28 | Zheng, W., Bellet, A., & Gallinari, P. (2018). **A distributed Frank–Wolfe framework for learning low-rank matrices with the trace norm.** Machine Learning, 107(8–10), 1457–1475. | Distributed Frank-Wolfe algorithm | 中心分布式。  结合了Naive DFW和Singular Vector Averaging策略。使得有Singular Vector Averaging的通信成本，和Naive DFW的收敛速率。 |

1. 概述

|  |  |
| --- | --- |
| Introduction or related works | 要点 |
| 分布式网络近年来得到了越来越多学者的关注。大体上，目前分布式网络可以分为两类。一类叫中心分布式，即整个分布式网络除了众多的节点外，还存在一个中心节点(center node)，其他节点都要和该中心节点通信；另一类叫完全分布式，也就是说，分布式网络中不存在中心节点，每一个节点都是平等的。其中一个节点，它周围的节点被叫做邻居节点(neighbor nodes)。在这种分布式网络中，通信发生在节点和它的邻居节点之间。  分布式网络不同于以往的单一节点网络。分布式网络具有以下特点：第一，针对大样本计算，分布式网络可以有效地减少单一节点的负荷，通过将计算成本分摊给每一个节点，来实现整体计算能力的提升。（通常的做法是将整体问题(global problem)分配成子问题(subproblem or distributed problem)，从而实现分布式计算）。如何设计子问题以减少计算成本，成为分布式研究的一个热点问题；第二，分布式网络通常会将数据(data)存储在当地(local)，避免了大量样本对单一节点的存储空间(storage)的高要求，而且，数据(data)存储在当地(local)通常是为了满足数据隐私的要求。第三，天下没有免费的午餐，单一节点的计算成本和存储空间要求都有所下降，那么节点之间交换信息势必会造成通信成本的上升。如何提高通信的效率，是分布式网络中可以深入研究的课题。具体来说，可以研究通信传输什么，比如，是传递parameter, gradient or partial data? 还可以研究通讯的方式，是中心式的还是完全式的？是否可以通过设定参数阈值条件，来判断此轮迭代中某节点是否参与通信，从而减少通信成本？等等。 | 分类：   1. 中心分布式网络； 2. 完全分布式网络。   分类依据：  是否有中心节点  特点：   1. 计算成本（软件方面） 2. 存储及数据隐私（硬件方面） 3. 通信成本（软件方面）   Open Problems and Future Directions：   1. 研究通信的效率，降低通信成本（算法层面） 2. 如何对子问题进行设计、改造、优化、近似，降低计算成本（算法层面） |

1. My Thoughts

目前所看论文，分布式终归是一个优化的方法。也就是说目前论文中，一个目标函数不用分布式也可以做，但是用了分布式，可以作为一个创新点（比如有分布式所带来的成本下降，算法更快）。常常是先定义出一个总体问题，再设计出子问题，子问题的结果可以收敛到原问题的结果。