

# 联邦学习算法研究

孙博

2020 年 10 月 29 日

## 目录

|   |           |
|---|-----------|
| <b>1 摘要</b>                                   | <b>5</b>  |
| <b>2 背景介绍</b>                                 | <b>5</b>  |
| <b>3 提高效率 and 效果</b>                          | <b>6</b>  |
| 3.1 挑战一-联邦学习中的非独立同分布数据                        | 6         |
| 3.1.1 解决non-IID Data 问题的策略                    | 7         |
| 3.2 联邦学习的优化算法                                 | 8         |
| 3.2.1 IID数据集的优化算法和收敛速度                        | 8         |
| <b>4 保护隐私</b>                                 | <b>9</b>  |
| 4.1 隐私保护技术                                    | 9         |
| 4.1.1 安全多方计算                                  | 9         |
| 4.1.2 同态加密                                    | 10        |
| 4.1.3 差分隐私                                    | 10        |
| <b>5 预备知识</b>                                 | <b>11</b> |
| 5.1 再生核希尔伯特空间Reproducing Kernel Hilbert Space | 11        |
| 5.1.1 常用的核函数                                  | 11        |
| 5.2 梯度下降                                      | 12        |
| 5.2.1 随机梯度下降Stochastic gradient descent       | 12        |
| 5.2.2 ADMM 交替向量乘子法                            | 12        |
| 5.3 张量分解                                      | 13        |
| 5.3.1 CP分解                                    | 13        |
| 5.3.2 Tucker 分解                               | 13        |
| 5.4 Dataset shift 数据集偏移                       | 13        |
| 5.4.1 有状态和无状态                                 | 13        |
| 5.5 激活函数                                      | 13        |
| 5.5.1 Sigmoid型函数                              | 13        |
| 5.5.2 整流线性单位函数(Rectified Linear Unit, ReLU)   | 14        |
| 5.5.3 Entropy熵                                | 16        |
| 5.5.4 Cross Entropy交叉熵                        | 16        |

|        |  |    |
|--------|--|----|
| 5.6    | 神经网络   | 16 |
| 5.6.1  | 感知器  | 17 |
| 5.6.2  | 分类问题总结   | 19 |
| 5.6.3  | 数据集  | 19 |
| 5.7    | 前馈神经网络   | 20 |
| 5.7.1  | 反向传播算法   | 21 |
| 5.7.2  | 优化问题   | 22 |
| 5.7.3  | 通用近似定理   | 22 |
| 5.8    | 循环神经网络   | 23 |
| 5.8.1  | 简单循环网络   | 23 |
| 5.8.2  | 循环神经网络的计算能力  | 23 |
| 5.8.3  | 简介实现   | 26 |
| 5.9    | 通过时间反向传播   | 27 |
| 5.10   | 语言模型   | 28 |
| 5.11   | 深度循环神经网络   | 28 |
| 5.11.1 | LSTM   | 29 |
| 5.11.2 | LSTM实现   | 30 |
| 5.11.3 | GRU  | 31 |
| 5.11.4 | 改进   | 31 |
| 5.12   | Attention  | 31 |
| 5.13   | Transformer  | 33 |
| 5.13.1 | 残差连接   | 34 |
| 5.13.2 | 机器翻译系统简介   | 35 |
| 5.14   | 基于梯度的方法的改进   | 35 |
| 5.14.1 | 动量法Momentum  | 35 |
| 5.14.2 | AdaGrad  | 37 |
| 5.14.3 | RMSProp  | 37 |
| 5.14.4 | AdaDelta   | 37 |
| 5.14.5 | Adam   | 38 |
| 5.15   | 建模   | 38 |
| 5.15.1 | RNN训练  | 40 |
| 6      | Private Empirical Risk Minimization                                      | 41 |
| 7      | CPFed: Communication-Efficient and Privacy-Preserving Federated Learning | 42 |
| 7.1    | 摘要   | 42 |
| 7.2    | 背景介绍   | 43 |
| 7.2.1  | 挑战   | 43 |
| 7.2.2  | 差分隐私DP   | 43 |
| 7.3    | 系统建模   | 44 |
| 7.3.1  | 联邦学习系统   | 44 |
| 7.3.2  | 威胁   | 44 |
| 7.3.3  | 目标   | 44 |

|        |  |    |
|--------|--|----|
| 7.4    | CPFed  | 44 |
| 7.4.1  | 提高通信效率-周期平均  | 44 |
| 7.4.2  | 防止隐私泄露-DP  | 45 |
| 7.4.3  | 提升模型精确度-安全集聚   | 45 |
| 7.4.4  | CPFed总体方案  | 46 |
| 7.5    | 隐私分析   | 47 |
| 7.5.1  | 收敛性分析  | 47 |
| 7.5.2  | 虚拟更新规则   | 48 |
| 7.6    | 实验   | 50 |
| 7.6.1  | 实验装置   | 50 |
| 7.6.2  | CPFED的收敛性  | 50 |
| 8      | Performance Analysis and Optimization in Privacy-Preserving Federated Learning | 52 |
| 8.1    | 背景   | 52 |
| 8.1.1  | 挑战   | 52 |
| 8.2    | 相关工作   | 52 |
| 8.3    | 预备知识   | 53 |
| 8.3.1  | 本地差分隐私   | 53 |
| 8.3.2  | 威胁模型-需应对的问题  | 55 |
| 8.4    | 隐私和收敛性分析   | 55 |
| 8.4.1  | Client-level DP 客户端级别DP  | 55 |
| 8.4.2  | Bound of the Moment  | 57 |
| 9.0.1  | 敏感度和隐私分析   | 57 |
| 9.1    | CDP中的折扣DISCOUNTING方法   | 59 |
| 9.1.1  | T的噪声重计   | 60 |
| 9.2    | 实验   | 60 |
| 9.2.1  | 结果评价   | 60 |
| 9.2.2  | 对CRD的评价  | 62 |
| 9.3    | 结论   | 62 |
| 9.4    | 公式推导   | 63 |
| 9.4.1  | Bound of the Moment 矩界   | 63 |
| 10.1   | 理论1的证明   | 65 |
| 10.2   | proof of Theorem 2   | 67 |
| 11     | Deep learning with differential privacy  | 69 |
| 11.1   | 差分隐私   | 70 |
| 11.1.1 | 组合理论 Composition Theorems  | 70 |
| 11.1.2 | 后记   | 71 |
| 11.2   | algorithmic foundations of DP笔记  | 71 |
| 11.2.1 | Final remarks on the definition  | 72 |

|  |           |
|--|-----------|
| <b>12 论文阅读: Federated Classification using Parsimonious Functions in Reproducing Kernel Hilbert Spaces</b> | <b>74</b> |
| 12.1 摘要与相关工作   | 74        |
| 12.1.1 相关工作  | 74        |
| 12.1.2 本文贡献  | 74        |
| 12.2 学习低复杂度RKHS表示  | 75        |
| 12.3 Federated Learning  | 76        |
| 12.4 Algorithm   | 76        |
| 12.5 Convergence of federated problem  | 76        |
| <b>13 图像处理实践—ALexNet网络</b>   | <b>76</b> |
| 13.1 架构  | 76        |
| 13.1.1 ReLu  | 76        |
| 13.1.2 局部响应归一化   | 76        |
| 13.1.3 重叠池化  | 76        |
| 13.1.4 降低过拟合   | 76        |
| 13.1.5 逐层归一化 (Layer-wise Normalization)  | 76        |
| 13.1.6 Dropout   | 77        |
| 13.2 实现  | 77        |
| 13.2.1 学习率的选择  | 79        |
| <b>14 联邦学习实践-Tensorflow federated</b>  | <b>79</b> |

# 联邦学习算法研究

孙博<sup>1</sup>

## 1 摘要

本文为文献 Advances and Open Problems in Federated Learning [1] 的笔记, 该文献介绍了联邦学习的最新进展, 相关研究方向.

联邦学习是一种机器学习架构, 在中央服务器或服务提供商的协调下, 多个实体(客户机)协作解决机器学习问题. 每个客户的原始数据存储在本地, 不进行交换或传输; 相反, 用于即时聚合的重点更新用于实现学习目标.

在这种设置中, 许多客户(例如移动设备或整个组织)在中央服务器(例如服务提供商)的协调下协作地训练模型, 同时保持训练数据分散. FL体现了集中数据收集和最小化的原则, 可以减轻由于传统的、集中的机器学习和数据科学方法所带来的许多系统隐私风险和成本. 在FL研究爆炸性增长的推动下, 本文讨论了近年来的进展, 并提出了大量的开放问题和挑战. 规范的联邦学习问题涉及到从存储在数以万计的远程设备上的数据学习单个全局统计模型. 我们的目标是在这样的约束下学习这个模型, 设备生成的数据在本地存储和处理, 只有中间更新与中央服务器定期通信. 目标可以用下列目标函数表示:

$$\min_w F(w), \text{ 其中 } F(w) := \sum_{k=1}^m p_k F_k(w).$$

这里  $m$  为设备总数,  $p_k \geq 0$  and  $\sum_k p_k = 1$ ,  $F_k$  是设备  $k$  的局部目标函数. 局部目标函数通常定义为局部数据的经验风险, 即  $F_k(w) = \frac{1}{n_k} \sum_{j_k=1}^{n_k} f_{j_k}(w; x_{j_k}, y_{j_k})$ , 其中  $n_k$  是本地可用的样本数量. 用户定义的术语  $p_k$  指定每个设备的相对影响, 两个原始设置分别是  $p_k = \frac{1}{n}$  或  $p_k = \frac{n_k}{n}$ , 其中  $n = \sum_k n_k$  是样本的总数.[2]

## 2 论文阅读: Federated Classification using Parsimonious Functions in Reproducing Kernel Hilbert Spaces

### 2.1 摘要与相关工作

**两个主要挑战** 统计挑战代理和用户相关联, 来自不同用户的数据是non-iid的  
系统挑战有时系统存储数据和通过网络有效传输数据的能力有限

#### 2.1.1 相关工作

- Federated learning: Collaborative machine learning without centralized training data[4] 在每个更新步骤中, 根据训练样本的数量计算每个代理的模型的加权平均值.
- Communication-efficient learning of deep networks from decentralized data[4] 使用梯度的加权平均来更新全局参数
- Adaptive Federated Learning in Resource Constrained Edge Computing Systems [37, 38] 使用基于梯度下降的方法训练的一类通用机器学习模型. 分析了分布梯度下降算法的收敛界, 并在此基础上

<sup>1</sup> 孙博, 2019级硕士研究生, 苏州蒙纳士联合研究生学院, 邮箱: ericsun20@qq.com.

提出了在给定资源预算条件下, 在局部更新和全局参数聚合之间寻找最佳平衡点以使损失函数最小化的控制算法.

- Privacy-preserving deep learning [23]为了进一步减少通信负载和增加隐私性, 使用了随机梯度下降变化, 它可以选择向服务器传输梯度的子集.

### 2.1.2 本文贡献

提出一种使用低复杂度再生核希尔伯特空间(RKHS)表示的算法, 每个代理在本地学习一个低复杂度的模型, 只将关键样本发送给中央服务器.

## 2.2 学习低复杂度RKHS表示

分类问题  $N$  feature-class 对  $(\mathbf{x}_n, y_n) \in \mathcal{T}$ . 特征  $\mathbf{X}_n \in \mathcal{X} \subset \mathbb{R}^p$  类  $y_n \in \{-1, +1\}$  是二元的. 目标学习近似函数  $f(x)$  满足  $f(x_n)$  与  $y_n$  的误差在一定的范围内, 即损失函数  $\ell(f(x), y) = 1 - \epsilon - yf(x)$ , 函数类  $\mathcal{C}$  和函数复杂度度量  $\rho(f)$  定义为如下优化问题

$$\begin{aligned} P = \min_{f \in \mathcal{C}} \quad & \rho(f) \\ \text{s. t.} \quad & \frac{1}{N} \ell(f(\mathbf{X}_n), y_n) \leq 0, \quad (\mathbf{X}_n, y_n) \in \mathcal{T}. \end{aligned} \quad (\text{PI})$$

$k(\mathbf{x}, \mathbf{s}; w)$  是核函数族, 其中  $\mathbf{x} \in \mathbb{R}^p$ ,  $\mathbf{s} \in \mathbb{R}^p$  是核中心,  $w \in \mathbb{R}$  是核参数. 且  $\mathcal{W} \subset \mathbb{R}$ ,  $\mathcal{S} \subseteq \mathbb{R}^p$ . 类函数为

$$\mathcal{C} = \left\{ f : f(\mathbf{x}) = \int_{\mathcal{S} \times \mathcal{W}} \alpha(\mathbf{s}, \mathbf{w}) k(\mathbf{x}, \mathbf{w}; w) d\mathbf{s} dw \right\}, \quad (1)$$

其中  $\alpha : \mathcal{S} \times \mathcal{W} \rightarrow \mathbb{R}$  是  $L_2(\mathcal{S} \times \mathcal{W})$  的系数函数(待求).

系数函数的稀疏度

$$\|\alpha\|_{L_0} = \int_{\mathcal{S} \times \mathcal{W}} \mathbb{I}[\alpha(\mathbf{s}, w) \neq 0] d\mathbf{s} dw. \quad (2)$$

ElasticNet在用Lasso回归太过(太多特征被稀疏为0), 而岭回归也正则化的不够(回归系数衰减太慢)的时候, 可以考虑使用ElasticNet回归来结合两者, 得到比较好的结果.

$$\begin{aligned} \rho(f) &= \frac{1}{2} \|\alpha\|_{L_2} + \gamma \|\alpha\|_{L_0} \\ &= \int_{\mathcal{S} \times \mathcal{W}} \frac{1}{2} \alpha^2(\mathbf{s}, w) + \gamma \mathbb{I}[\alpha(\mathbf{s}, w) \neq 0] d\mathbf{s} dw. \end{aligned} \quad (3)$$

原则上, 我们希望尽可能的大来支持稀疏性. 然而在实践中, 我们要减少价值受益于正则化的效果(见注3).

再生核希尔伯特空间理论的低复杂度分类问题由类函数  $\mathcal{C}$  和函数复杂度定义.

$$\begin{aligned} P = \min_{\alpha} \quad & \int_{\mathcal{S} \times \mathcal{W}} \frac{1}{2} \alpha^2(\mathbf{s}, w) + \gamma \mathbb{I}[\alpha(\mathbf{s}, w) \neq 0] d\mathbf{s} dw \\ \text{s. t.} \quad & \frac{1}{N} \ell(f(\mathbf{x}_n), y_n) \leq 0, \quad (\mathbf{x}_n, y_n) \in \mathcal{T} \\ & f(\mathbf{x}) = \int_{\mathcal{S} \times \mathcal{W}} \alpha(\mathbf{s}, w) k(\mathbf{x}, \mathbf{s}; w) d\mathbf{s} dw, \end{aligned} \quad (\text{PC})$$

问题(PC)把(PI)的函数  $f$  的搜索换为系数  $\alpha$  的搜索. 两者是 **otherwise** 等价的.

**Theorem 2.1.** 如果满足以下条件, 则问题(PC)具有强对偶性: (i)核 $k(\cdot, \mathbf{s}; w)$ 不具有点质量, 且 (ii)满足Slater条件.

**Remark 1.** 问题(PC)适合使用约束的分类函数, 而不是正则化最小化问题. 使用约束问题的好处有两点:它允许在对偶域解决问题, 对偶域的解决为我们的学习问题提供了关键样本的信息.

**Remark 2.** 函数类 $\mathcal{C}$ 与RKHS密切相关. 对于(66)中的固定核参数 $w$ ,  $f(\mathbf{x}) = \int_{\mathcal{S}} \alpha(\mathbf{s}, w) k(\mathbf{x}, \mathbf{s}; w) d\mathbf{s}$ 是内核 $k$ 生成的RKHS的整数表示. [11, 39, 40]. 与更传统的级数表示相比, 这种积分表示实际上是一种简化. 如果由级数表示, 很难找到核的稀疏集. 但(PC)中的问题在对偶域中是可以处理的. 然后, 我们还可以将核心参数 $w$ 纳入搜索空间. 这样不仅优化了内核的位置和宽度, 甚至可以在RKHS的联合中表示, 即表示混合了不同宽度的核[11].

## 2.3 Federated Learning

## 2.4 Algorithm

## 2.5 Convergence of federated problem

## 参考文献

- [1] P. Kairouz, H. B. McMahan, K. Bonawitz, Z. Charles, R. G. L. D. Oliveira, Z. Garrett, P. B. Gibbons, A. Ozg, W. Song, S. U. Stich, and F. X. Yu, “Advances and open problems in federated learning,” Tech. Rep., 2019.
- [2] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” *arXiv preprint arXiv:1908.07873*, 2019.
- [3] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated machine learning: Concept and applications,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.
- [4] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” feb 2016. [Online]. Available: <http://arxiv.org/abs/1602.05629>
- [5] K. Hsieh, A. Phanishayee, O. Mutlu, and P. B. Gibbons, “The non-iid data quagmire of decentralized machine learning,” *arXiv preprint arXiv:1910.00189*, 2019.
- [6] J. Mancuso, “Privacy-preserving machine learning 2018: a year in review,” 2018.
- [7] A. C. Yao, “Protocols for secure computations,” *23rd annual symposium on foundations of computer science*, pp. 160–164, 1982.
- [8] A. C. Yao, “How to generate and exchange secrets,” in *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*, 1986, pp. 162–167.
- [9] C. Dwork, “Differential privacy: A survey of results,” in *Proceedings of the 5th International Conference on Theory and Applications of Models of Computation*, ser. TAMC’ 08. Berlin, Heidelberg: Springer-Verlag, 2008, p. 1–19.

- [10] S. Fletcher and M. Z. Islam, “Differentially private random decision forests using smooth sensitivity,” *Expert Systems with Applications*, vol. 78, pp. 16 – 31, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417417300428>
- [11] M. Peifer, L. F. O. Chamon, S. Paternain, and A. Ribeiro, “Sparse Multiresolution Representations With Adaptive Kernels,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 2031–2044, 2020.
- [12] 苏剑林, “训练集、验证集和测试集的意义,” [Blog post], Oct 14, 2017, retrieved from <https://kexue.fm/archives/4638>.
- [13] J. L. Elman, “Finding structure in time,” *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.
- [14] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, pp. 1735–80, 12 1997.
- [15] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *Computer Science*, 2014.
- [16] S. P. Kasiviswanathan, M. Rudelson, and A. Smith, “The power of linear reconstruction attacks,” 2012.
- [17] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, “Local privacy and statistical minimax rates,” in *Proceedings of the 54th Annual Symposium on the Foundations of Computer Science (FOCS '13)*, Berkeley, CA, USA, October 26–29 2013. [Online]. Available: <http://dx.doi.org/10.1109/FOCS.2013.53>
- [18] K. Fukuchi, Q. K. Tran, and J. Sakuma, “Differentially private empirical risk minimization with input perturbation,” ArXiv, Tech. Rep. arXiv:1710.07425 [stat.ML], October 2017. [Online]. Available: <https://arxiv.org/abs/1710.07425>
- [19] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate, “Differentially private empirical risk minimization,” *Journal of Machine Learning Research*, vol. 12, pp. 1069–1109, March 2011. [Online]. Available: <http://jmlr.csail.mit.edu/papers/v12/chaudhuri11a.html>
- [20] J. Zhang, Z. Zhang, X. Xiao, Y. Yang, and M. Winslett, “Functional mechanism: Regression analysis under differential privacy,” *Proceedings of the VLDB Endowment*, vol. 5, no. 11, pp. 1364–1375, July 2012. [Online]. Available: <http://dx.doi.org/10.14778/2350229.2350253>
- [21] B. I. P. Rubinstein, P. L. Bartlett, L. Huang, and N. Taft, “Learning in a large function space: Privacy-preserving mechanisms for SVM learning,” *Journal of Privacy and Confidentiality*, vol. 4, no. 1, pp. 65–100, 2012. [Online]. Available: <http://repository.cmu.edu/jpc/vol4/iss1/4/>
- [22] R. Hu, Y. Gong, and Y. Guo, “Cpfed: Communication-efficient and privacy-preserving federated learning,” *arXiv preprint arXiv:2003.13761*, 2020.
- [23] R. Shokri and V. Shmatikov, “Privacy-preserving deep learning,” in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, Oct. 2015, pp. 1310–1321.
- [24] L. Zhao, Q. Wang, Q. Zou, Y. Zhang, and Y. Chen, “Privacy-preserving collaborative deep learning with unreliable participants,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1486–1500, to appear 2020.



- [25] J. Lee and D. Kifer, “Concentrated differentially private gradient descent with adaptive per-iteration privacy budget,” in *Proc. ACM SIGKDD*, New York, NY, USA, 2018, pp. 1656–1665.
- [26] C. Li, P. Zhou, L. Xiong, Q. Wang, and T. Wang, “Differentially private distributed online learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 8, pp. 1440–1453, Aug 2018.
- [27] R. C. Geyer, T. Klein, and M. Nabi, “Differentially Private Federated Learning: A Client Level Perspective,” *arXiv*, 2017. [Online]. Available: <http://arxiv.org/abs/1712.07557>
- [28] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, “A Hybrid Approach to Privacy-Preserving Federated Learning,” in *Proc. ACM Workshop on AISec*, London, UK, 2019, pp. 1–11.
- [29] T. Li, Z. Liu, V. Sekar, and V. Smith, “Privacy for Free: Communication-Efficient Learning with Differential Privacy Using Sketches,” *arXiv*, 2019. [Online]. Available: <https://arxiv.org/abs/1911.00972>
- [30] C. Dwork, F. McSherry, K. Nissim, and A. Smith, *Calibrating Noise to Sensitivity in Private Data Analysis*, jan 2006, vol. Vol. 3876.
- [31] H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang, “Learning differentially private recurrent language models,” *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, pp. 1–14, 2018.
- [32] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, “Deep Learning with Differential Privacy,” *arXiv e-prints*, p. arXiv:1607.00133, Jul. 2016.
- [33] F. D. McSherry, “Privacy integrated queries: an extensible platform for privacy-preserving data analysis,” in *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, 2009, pp. 19–30.
- [34] C. Dwork, G. N. Rothblum, and S. Vadhan, “Boosting and differential privacy,” in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, 2010, pp. 51–60.
- [35] C. Dwork and A. Roth, “The Algorithmic Foundations of Differential Privacy,” *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2014.
- [36] M. Arapinis, D. Figueira, and M. Gaboardi, “Sensitivity of counting queries,” *Leibniz International Proceedings in Informatics, LIPIcs*, vol. 55, no. 120, pp. 1–13, 2016.
- [37] S. Wang and Others, “Adaptive Federated Learning in Resource Constrained Edge Computing Systems,” *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205–1221, jun 2019.
- [38] H. Yu, S. Yang, and S. Zhu, “Parallel restarted sgd with faster convergence and less communication: Demystifying why model averaging works for deep learning,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 5693–5700.
- [39] M. Peifer, L. F. O. Chamon, S. Paternain, and A. Ribeiro, “Locally Adaptive Kernel Estimation Using Sparse Functional Programming,” in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 2022–2026.

- [40] —, “Sparse Learning of Parsimonious Reproducing Kernel Hilbert Space Models,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 3292–3296.