A Collaborative Training Algorithm for Distributed Learning

Joel B. Predd, Sanjeev R. Kulkarni and H. Vincent Poor
IEEE Trans. on Info. Theory, April 2009

Presentation by: Venugopalakrishna Y. R., SPC Lab, IISc

7^{th} April, 2012

Outline

- Spatial Field Estimation
- Preliminaries of Supervised Learning
- Model considered
- Centralized Kernel-linear least square estimator
- Relaxed problem to facilitate distributed learning
- Successive Projection Algorithm
- Proposed Algorithm

Spatial Field Estimation

- Estimation of a scalar field by employing WSNs, specifically, spatial distribution of power in a given frequency band
- Nodes measure the field at their location and can transmit to a central node, where field can be estimated
- Nodes are characterized by energy and bandwidth constraints, hence local communication is preferred
- Local message passing algorithms for distributed inference in WSNs where each node or set of anchor nodes learn the global spatial field
- Our Interest: To learn the local field well

Supervised Learning

- Given input features, RV $X \in \mathcal{X}$ and corresponding outcome measurements, RV $Y \in \mathcal{Y}$, we need to learn the function $f : \mathcal{X} \to \mathcal{Y}$ that minimizes a loss function
- If loss function is MSE, Conditional mean estimator is the solution. But, we are required to know the joint PDF P_{XY}
- Non-parametric least squares approach is considered
- One approach: Regularized Kernel methods

Regularized Kernel Methods

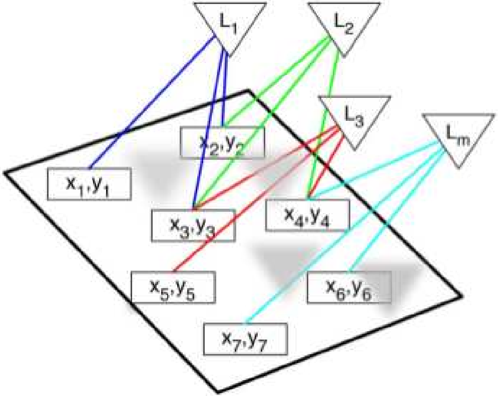
- $\mathcal{H}_k$ - RKHS induced by a positive definite kernel $K(.,.) : \mathcal{X} \times \mathcal{X} \to \mathcal{R}$
- Optimization problem

$$\min_{f \in \mathcal{H}_k} \left[\frac{1}{n} \sum_{i=1}^{n} (f(x_i) - y_i)^2 + \lambda ||f||^2_{\mathcal{H}_k} \right] \quad (1)$$

- Representer Theorem: Let $g_n \in \mathcal{H}_k$ be the minimizer of (1). Then there exists $\mathbf{c_n} \in \mathcal{R}^n$ s.t.

$$g_n = \sum_{i=1}^{n} c_{n,i} K(., x_i) \quad (2)$$

Model for Distributed Learning



$S_n = \{1, \ldots, n\}$ - set of n nodes (n training examples), m - learning agents, $S_n^j \subseteq S_n$ - set of nodes interacting with j^{th} learning agent

Altering the optimization problem

$$\min_{f \in \mathcal{H}_k} \left[\frac{1}{n} \sum_{i=1}^{n} (f(x_i) - y_i)^2 + \lambda ||f||^2_{\mathcal{H}_k} \right] \tag{3}$$

- Above problem requires training examples at a central node
- Introduce a decision rule $f_j \in \mathcal{H}_k$ for each agent $j = 1 \dots m$

$$\begin{aligned} \min \left[\sum_{i=1}^{n} (z_i - y_i)^2 + \sum_{j=1}^{m} ||f_j||^2_{\mathcal{H}_k} \right] s.t. \\ z_i = f_j(x_i) \forall i \in S_n, j = 1 \dots m \\ f_j \in \mathcal{H}_k, j = 1 \dots m \end{aligned} \tag{4}$$

- If $\frac{1}{n} \sum \lambda_j = \lambda$, then $g_n^j = g_n$, where g_n is solution to (3) and $(\mathbf{z}, g_n^1, \dots, g_n^m)$ is solution to (4)

Further Relaxation

- Previous problem: Centralized Regression and Global agreement
- This motivates distributed regression and local agreement

$$\min \left[\sum_{i=1}^{n} (z_i - y_i)^2 + \sum_{j=1}^{m} ||f_j||^2_{\mathcal{H}_k} \right] s.t. \tag{5}$$

$$z_i = f_j(x_i) \forall i \in S_n^j, j = 1 \ldots m$$

$$f_j \in \mathcal{H}_k, j = 1 \ldots m$$

Relaxed problem

- Let $\mathcal{H} = \mathcal{R}^n \times \mathcal{H}_k$ be the Hilbert space with norm $||(\mathbf{z}), f_1, \ldots, f_m)||^2 = ||\mathbf{z}||_2^2 + \sum_{j=1}^m \lambda_j ||f_j||_{\mathcal{H}_k}^2$
- Interpretation: (5)is orthogonal projection of $(y, 0, \ldots, 0) \in \mathcal{H}$ on to the set $C = \bigcap_{j=1}^m C_j \subset \mathcal{H}$ with

$$C_j = \{(\mathbf{z}, f_1, \ldots, f_m) : f_j(x_i) = z_i \;\; \forall i \in S_n^j, \ddagger \in \mathcal{R}^n\}$$

- Successive Orthogonal Projection Algorithm can be used

Successive Orthogonal Projection

- Simple algorithm to compute orthogonal projection of a point to the intersection of convex sets, using a sequence of projections on to the sets
- Slides of Stephen Boyd

Projection on to C_j

- Projecting a point $v = (\mathbf{z}, f_1, \ldots, f_m)$ on to C_j results in $P_{C_j}(v) = (\mathbf{z}^*, f_1^*, \ldots, f_m^*)$ where

$$\begin{aligned} f_k^* &= f_k && \forall k \neq j \\ f_j^* = \arg\min_{f \in \mathcal{H}_k} \sum_{i \in S_n^j} (f(x_i) - z_i)^2 &+ \lambda_j ||f - f_j||^2_{\mathcal{H}_k} && \\ z_i^* &= z_i && \forall i \notin S_n^j \\ z_i^* &= f_j^*(x_i) && \forall i \in S_n^j \end{aligned}$$

Some Results

- As $T \to \infty$, $g_n^{j,T} = g_n^j$
- if graph is connected then g_n^j converges to global estimate g_n

Thank you