Gossip Algorithm

Yilin Mo

July 2, 2015

We model a network composed of n agents as a graph $G = \{V, E\}$. $V = \{1, 2, \ldots, n\}$ is the set of vertices representing the agents. $E \subseteq V \times V$ is the set of edges. $(i, j) \in E$ if and only if sensor i and j can communicate directly with each other. We will always assume that G is undirected, i.e. $(i, j) \in E$ if and only if $(j, i) \in E$. We further assume that there is no self loop, i.e., $(i, i) \notin E$.

At each time step, we assume a pair of nodes (i, j) is randomly selected with probability p_{ij}, where $(i, j) \in E$. They communicate with each other and perform the following update:

$$x_i(k+1) = x_j(k+1) = (x_i(k) + x_j(k))/2.$$

All the other nodes that are not selected keep their own states:

$$x_l(k+1) = x_l(k), \forall l \notin \{i, j\}.$$

Let us define matrix

$$W_{ij} = I - (e_i - e_j)(e_i - e_j)^T/2.$$

Then the update equation can be written in matrix form as

$$x(k+1) = W(k)x(k) = W_{ij}x(k), \text{ with probability } p_{ij}.$$

Define $y(k) = x(k) - x_{ave} = (I - J)x(k)$ (why?).

$$y(k+1) = (W(k) - J)y(k).$$

Problem:

- Under what condition does $x(k)$ converges to the $x_{ave} = Jx(0)$, where $J = \mathbf{1}\mathbf{1}^T/n$?
- How fast is the convergence rate?

1 Products of Random Numbers

Let us assume that $\alpha(k)$ are i.i.d. distributed and define

$$\beta(k) = \alpha(k) \ldots \alpha(0).$$

1. $\beta(k)$ converges to 0 almost surely if

$$P(\lim_{k\to\infty} \beta(k) = 0) = 1.$$

2. $\beta(k)$ converges to 0 in L_p if

$$\lim_{k\to\infty} \mathbb{E}\beta(k)^p = 0.$$

The L_p convergence rate can be defined as

$$\rho_p = \sqrt[k]{\mathbb{E}\left(\beta(k)^p\right)}$$

A special case is convergence in mean square sense:

$$\lim_{k\to\infty} \mathbb{E}\beta(k)^2 = 0$$

Theorem 1. *$\beta(k)$ converges in L_p if and only if*

$$\rho_p = \mathbb{E}\left(\alpha(0)^p\right) < 1.$$

Furthermore,

$$P\left(\lim_{k\to\infty} \sqrt[k]{\beta(k)} = \exp\left[\mathbb{E}\left(\log \alpha(k)\right)\right]\right) = 1.$$

Proof. Since $\alpha(k)$ are independent of each other,

$$\mathbb{E}\beta(k)^p = \prod_{t=0}^{k} \mathbb{E}\alpha(t)^p.$$

Hence, $\rho_p = \mathbb{E}\alpha(t)^p$.

On the other hand,

$$\log \beta(k) = \sum_{t=0}^{k} \log \alpha(t).$$

By LLN,

$$\lim_{k\to\infty} \frac{\log \beta(k)}{k} = \mathbb{E}\left(\log \alpha(t)\right), \text{ almost surely}$$

□

Let us assume that

$$\alpha(k) = \begin{cases} 0 & \text{with probability } 0.5 \\ 2 & \text{with probability } 0.5 \end{cases}$$

Then $\beta(k)$ converges almost surely, but does not converge in L_p.

2 Products of Random Matrices

The following theorem can be seen as a generalization of LLN for noncommutative case

Theorem 2. *Let $X(k)$ be i.i.d. distributed, if*

$$\mathbb{E}\left[\max(\log\|X(k)\|, 0)\right] < \infty,$$

then

$$\lim_{k\to\infty} k^{-1}\log\|X(k)X(k-1)\dots X(1)\| = \rho_{as},$$

with probability 1, where ρ_{as} is defined as

$$\rho_{as} = \lim_{k\to\infty} k^{-1}\mathbb{E}\left(\log\|X(k)X(k-1)\dots X(1)\|\right).$$

For general cases, the almost surely convergence rate ρ_{as} is still unknown.

Now let us consider the mean square convergence rate. First look at the W_{ij} matrix, we know that

- W_{ij} is symmetric.
- $W_{ij} \le I$.
- $W_{ij}\mathbf{1} = \mathbf{1}$.

Define

$$W \triangleq \mathbb{E}W(k) = \sum_{(i,j)\in E} p_{ij}W_{ij}.$$

Therefore,

- W is symmetric.
- $W \le I$.
- $W\mathbf{1} = \mathbf{1}$.

Define $\mathcal{W} = W - J$, then

$$y(k+1) = (W(k) - J)y(k) \implies \mathbb{E}y(k+1) = \mathcal{W}\mathbb{E}y(k).$$

Therefore,

$$\mathbb{E}y(k) = \mathcal{W}^k y(0).$$

By Jensen's inequality,

$$\mathbb{E}\|y(k)\|^2 \ge \|\mathbb{E}y(k)\|^2 = \|\mathcal{W}^k y(0)\|^2. \tag{1}$$

Thus, there exists an $y(0)$ (which is the eigenvector of $\lambda_2(W)$, such that

$$\mathbb{E}\|y(k)\|^2 \ge \lambda_2(W)^{2k}\|y(0)\|^2$$

On the other hand,

$$y(k+1)^T y(k+1) = y(k)^T (W(k) - J)^2 y(k) = y(k)^T (W(k) - J) y(k).$$

Hence,

$$\mathbb{E}\|y(k+1)\|^2 = \mathbb{E}\left[y(k)^T \mathbb{E}((W(k) - J)|y(k)) y(k)\right] = \mathbb{E}\left[y(k)^T \mathcal{W} y(k)\right] \leq \lambda_2(W) \mathbb{E}\|y(k)\|^2.$$

Thus,

$$\mathbb{E}\|y(k)\|^2 \leq \lambda_2(W)^k \|y(0)\|^2.$$

Hence, the mean square convergence rate satisfies

$$\lambda_2(W)^2 \leq \rho_2 \leq \lambda_2(W).$$

The gossip algorithm converges in the mean square sense if and only if $\lambda_2(W) < 1$.

To get the exact mean square convergence rate, we need to consider

$$Y(k) = y(k) y(k)^T.$$

Clearly,

$$Y(k+1) = (W(k) - J) Y(k) (W(k) - J).$$

Hence,

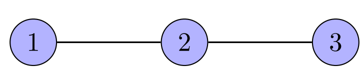
$$\mathbb{E} Y(k+1) = \sum_{(i,j) \in E} p_{ij} (W_{ij} - J) \mathbb{E} Y(k) (W_{ij} - J).$$

If we define the operator $\mathcal{A} : \mathbb{S}_n \to \mathbb{S}_n$:

$$\mathcal{A}(X) = \sum_{(i,j) \in E} p_{ij} (W_{ij} - J) X (W_{ij} - J).$$

$\mathcal{A}(X)$ is a linear operator on $\mathbb{S}_n$. The mean square convergence rate ρ_2 will be the operator norm of $\mathcal{A}$.

Examples:



Assume that $p_{12} = 0.5$, $p_{23} = 0.5$. Hence,

$$W = \frac{1}{4} \begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}.$$

The eigenvalues are $1, 0.75, 0.25$. The corresponding eigenvectors are

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, v_3 = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}.$$

Hence, we know that $9/16 \le \rho_2 \le 3/4$. Now let us define

$$v_a = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, v_b = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

We argue that $y(k)$ is either $\theta(k)v_a$ or $\theta(k)v_b$. (Why?)

If $y(k) = \theta(k)v_a$, then if $(1,2)$ is selected, $y(k+1) = y(k)$. On the other hand, if $(2,3)$ is selection, then

$$y(k+1) = \theta(k) \begin{bmatrix} 1 \\ -0.5 \\ -0.5 \end{bmatrix} = -\frac{\theta(k)}{2} v_b.$$

Hence

$$\|y(k+1)\|^2 = \begin{cases} \|y(k)\|^2 & \text{with probability } 0.5 \\ 0.25\|y(k)\|^2 & \text{with probability } 0.5 \end{cases}$$

The true mean square convergence rate should be

$$\rho_2 = 0.5 \times 1 + 0.5 \times 0.25 = \frac{5}{8}.$$