SOME RANDOM, DISTORTED GOSSIP

Vivek Borkar

TIFR

September 8, 2014

Abstract

Gossip has always been a thrust area on academic campuses. Taking advantage of my greying hair and aging brain cells, I shall indulge in that favourite pastime of academic dotards - some random, distorted gossip.

To add a veneer of culture (of sorts), I shall draw occasional parallels with thoughts expressed by some leading poet-philosophers of the last century.

Outline

Classical (‘*vanilla*’) gossip

Outline

Classical ('*vanilla*') gossip

Random gossip

Nonlinear gossip

Optimal gossip

Outline

Classical (‘*vanilla*’) gossip

Random gossip

Nonlinear gossip

Optimal gossip

(Modi-fied gossip)

Thoughts of poet-philosophers will appear in red.

CLASSICAL GOSSIP

‘Gossip’ algorithm

$$x_i(n+1) = \sum_{j=1}^{d} p(j|i)x_j(n), \; n \geq 0.$$

$P = [[p(j|i)]]_{1 \leq i,j \leq d}$ irreducible stochastic matrix with unique stationary distribution $\pi \Longrightarrow x(n) \rightarrow \pi^T x(0)\mathbf{1}$.

Research focus on rate of convergence: Design a ‘good’ P ((doubly) stochastic, low |second eigenvalue|, $\cdots$) (Boyd, Shah, Ghosh, $\cdots$)

Ref: ‘*Gossip Algorithms*’, D. Shah, NOW Publishers, 2009.

Often a component of a ‘larger’ scheme:

$$x_i(n+1) = (1-a)x_i(n) + a\sum_{j=1}^{d} p(j|i)x_j(n) + \cdots, \ n \geq 0.$$

Examples: Distributed computation, Synchronization, ‘Flocking’, Coordination of mobile agents

The objective often is ‘*consensus*’.

Well, I try my best to be just like I am,
but everybody wants you to be just like them.

(Bob Dylan)

The DeGroot model

Models opinion formation in society.

$$x_i(n+1) = (1-a)x_i(n) + a\sum_{j=1}^{d} p(j|i)x_j(n), \; n \geq 0.$$

New opinion a convex combination of own previous opinion and opinions of neighbors/peers/friends.

Convergence $\Longrightarrow$ asymptotic agreement.

I get by with a little help from my friends.

(Lennon-McCartney)

The DeGroot model

Models opinion formation in society.

$$x_i(n+1) = (1-a)x_i(n) + a\sum_{j=1}^{d} p(j|i)x_j(n), \;\; n \geq 0.$$

New opinion a convex combination of own previous opinion and opinions of neighbors/peers/friends.

Convergence $\Longrightarrow$ asymptotic agreement.

I get by with a little help from my friends,
I get high with a little help from my friends.

(Lennon-McCartney)

RANDOM GOSSIP

What about **random** gossip?

$$x_i(n+1) = (1-a)x_i(n) + a x_{\xi_{n+1}(i)}(n),$$

where $\xi_n(i)$ IID $\approx p(\cdot|i)$.

Convergence?

What about **random** gossip?

$$x_i(n+1) = (1-a)x_i(n) + a x_{\xi_{n+1}(i)}(n),$$

where $\xi_n(i)$ IID $\approx p(\cdot|i)$.

Convergence?

Yes!! And **consensus:** $x(n) \to c\mathbf{1}$, but c may not be $\pi^T x(0)$!

Jaatey the Japaan pahoonch gaye Cheen, samaz gaye na?

(Majrooh Sultanpuri)

Analysis based on re-writing the iteration as

$$x_i(n+1) = (1-a)x_i(n) + a\sum_{j=1}^{d} p(j|i)x_j(n) + aM_j(n+1),$$

where $\{M(n)\}$ is a martingale difference sequence. This is a '*constant step-size stochastic approximation*'.

Fact: Standard 'intuition' would suggest asymptotically a random walk along the degenerate direction $c\mathbf{1}, c \in \mathbf{R}$, but we still get convergence because 'noise' $\{M(n)\}$ is also killed asymptotically at a fast enough rate.

But what if we want the actual average $\pi^T x(0)$?

Alternative scheme based on the 'Poisson equation': for $f(i) = x(0)$,

$$V(i) = f(i) - \beta + \sum_j p(j|i)V(j), \ 1 \leq j \leq d. \qquad (1)$$

Solution $(V(\cdot), \beta)$ satisfies: β unique, $= \pi^T f$, V unique up to additive scalar.

Can solve (1) by the 'relative value iteration'

$$V^{n+1}(i) = f(i) - V^n(i_0) + \sum_j p(j|i)V^n(j), \ n \geq 0.$$

The 'offset' $V^n(i_0)$ stabilizes the iteration, other choices are possible (e.g., $\frac{1}{d}\sum_k V^n(k)$).

'*Reinforcement learning*': stochastic approximation version of RVI – for a simulated chain $\{X_n\} \approx p(\cdot|\cdot)$.

$$V^{n+1}(i) = (1 - a(n)I\{X_n = i\})V^n(i) + a(n)I\{X_n = i\}(k(i) - V^n(i_0) + V^n(X_{n+1})).$$

Then $V^n(i_0) \to \beta$ a.s. (**Not** fully decentralized: needs $V^n(i_0)$ to be broadcast. Can replace it by $\frac{1}{d}\sum_k V^n(k)$ which can be calculated in a distributed manner by another gossip on a faster time scale.)

With every mistake we must surely be learning

(George Harrison)

NONLINEAR GOSSIP

'Multiplicative' analog of the previous case: for $f(i) > 0$, choose $V^0(i) > 0 \ \forall \ i$ and do:

$$V^{n+1}(i) = \frac{f(i) \sum_j p(j|i) V^n(j)}{V^n(i_0)}, \ n \geq 0.$$

More generally, for irreducible nonnegative $Q = [[q(i,j)]]$, set

$$f(i) = \sum_k q(i,k), \ p(j|i) = \frac{q(i,j)}{q(i)}.$$

Then $V^n(i_0) \rightarrow$ the Perron-Frobenius eigenvalue of Q, $V^n \rightarrow$ the corresponding eigenvector.

('power' method)

Applications : ranking, risk-sensitive control

‘Learning’ version: for $V^0(\cdot) > 0$,

$$\begin{aligned} V^{n+1}(i) \;=\; & (1 - a(n)I\{X_n = i\})V^n(i) + \\ & a(n)I\{X_n = i\}\left(\frac{f(i)V^n(X_{n+1})}{V^n(i_0)}\right). \end{aligned}$$

Numerically better even when the eigenvalue is known!

(The first term on RHS scales slower than the second.)

Similar evolution occurs in models of emergent networks (Jain - Krishna)

For tracking in slowly varying environment:

Decreasing stepsize $\Longrightarrow$

learning eventually slower than environment $\Longrightarrow$

cannot track $\Longrightarrow$ use constant stepsize $a(n) \equiv a > 0$.

And the first one now will later be last,

For the times, they are a-changin'

(Bob Dylan)

Tsitsiklis model: Distributed stochastic approximation (e.g., stochastic gradient scheme) with consensus objective.

$$x_i(n+1) = \sum_{j=1}^{N} p(j|i)x_j(n) + a(n)[h_i(x(n)) + M_i(n+1)].$$

$P = [[p(j|i)]]$ irreducible stochastic matrix.

Standard paradigm for network-based computation, particularly optimization, e.g., over sensor networks

Come together right now, over me.

(Lennon-McCartney)

A ‘quasi’-linear version:

$$x_i(n+1) = \sum_j p_{x(n)}(j|i)x_j(n) + a(n)\left[h_i(x(n)) + M_i(n+1)\right]$$

$P_x := [[p_x(j|i)]]$ stochastic irreducible with unique stationary distribution π_x.

Asymptotically, decouples into identical trajectories of

$$\dot{y}(t) = \sum_i \pi_{y(t)\mathbf{1}}(i)h_i(y(t)).$$

Jahaan bhi le jaaye raahen, hum sang hain.

(Shailendra)

Application: 'leaderless swarm optimization' by a swarm of n agents.

Let $N(i) :=$ { neighbors of $i\}$ with $|N(i)| = n$.
Also, $i \in N(j) \Longleftrightarrow j \in N(i)$. Let for $T > 0$,

$$
\begin{aligned}
h(x) &= -\nabla F(x), \\
p_x(j|i) &= \frac{1}{n} e^{-(F(x_j)-F(x_i))^+/T}, \;\; j \in N(i), \\
&= 0, \qquad\qquad j \notin N(i) \cup \{i\}, \\
&= 1 - \sum_{k \in N(i)} p(k|i), \;\; j = i.
\end{aligned}
$$

Then $\pi_x(i) = \frac{e^{-\frac{F(x_i)}{T}}}{\sum_k e^{-\frac{F(x_k)}{T}}}$ which concentrates on the 'best' j.

Empirical experiments on standard functions (Rastrigin, Rosenbrock, Griewank, Shwefel) show much better result compared to the single agent case (global minimum, or at least a very good local minimum).

Ek akela thak jaayega, milkar kadam badhana.

(Sahir Ludhianvi)

Fully nonlinear case:

$$x_i(n+1) = F_i(x(n)) + a(n)\left[h_i(x(n)) + M_i(n+1)\right].$$

Suppose:

$F^{(n)} := F \circ F \circ \cdots \circ F(n \text{ times}) \stackrel{n\uparrow\infty}{\rightarrow} \check{F}$,

$\mathcal{I} :=$ the invariant set of $\check{F}$.

Then under suitable technical hypotheses, limiting dynamics $\approx$ the o.d.e. restricted to $\mathcal{I}$.

Application: $x(n+1) = F(x(n))$ can be a recursive scheme for calculating projection to a constraint set $\mathcal{C}$. Then $\mathcal{I} = \mathcal{C}$.

Examples: 1. alternating projections for $\mathcal{C} =$ intersection of subspaces

2. Iusem-De Pierro variant of the Boyle-Dykstra-Han algorithm for $\mathcal{C} =$ intersection of convex sets

3. Optimization on matrix manifolds????

OPTIMAL GOSSIP

Gossip for opinion manipulation (e.g., advertising):

P_1 := submatrix of P corresponding to $n - m$ rows and corresponding columns,
P_2 := submatrix of P corresponding to the same $n - m$ rows and remaining m columns.

These m columns correspond to nodes whose 'opinion' is frozen at x^*. Then we have (in $\mathbf{R}^{n-m}$):

$$x(n+1) = x(n) + a(n)\left[P_1 x(n) + P_2 x^* \mathbf{1}\right].$$

Assume P_1 strictly sub-stochastic, irreducible. Then:

$x(n) \to x^* \mathbf{1}$ exponentially at rate $\lambda :=$ the Perron-Frobenius eigenvalue of P_1.

$\Longrightarrow$ consensus on a pre-specified value.

Jo tumko ho pasand wohi baat karenge.

Tum din ko agar raat kaho raat kahenge.

(Indeevar)

Objective: Minimize λ over all subsets of cardinality m (i.e., find the m most important nodes for information dissemination)

Hard combinatorial problem, even the nonlinear programming relaxation is highly non-convex and the projected gradient scheme with multi-start does not do too well.

Christ, you know it ain't easy,

you know how hard it can be.

(Lennon)

$\Longrightarrow$ Use '**engineer's licence**':

For $\tau :=$ the first passage time to frozen nodes, $\lambda = -\lim_{t\uparrow\infty} \frac{1}{t} \log P(\tau > t)$ and $E[\tau] = \sum_{t=0}^{\infty} P(\tau \geq t)$.

$\Longrightarrow$ Use $E[\tau]$ as a surrogate cost.

This is *monotone and supermodular* $\Longrightarrow$ greedy scheme is $\left(1 - \frac{1}{e}\right)$-optimal (Nemhauser-Wolsey-Fisher)

Important observation: best m nodes $\neq$ top m nodes according to individual merit!

What about controlling the transition probabilities?

Consider controlling the nonlinear o.d.e.

$$\dot{x}(t) = \alpha(P_1^{u(t)} - I)x(t) + \alpha P_2^{u(t)}(x^*\mathbf{1}) + (1-\alpha)F(x(t))$$

with ‘cost’

$$E\left[\int_0^\infty e^{-\beta t}\sum_i |x_i(t) - x^*|^2 dt\right].$$

Here $P_\cdot^u = [[p(j|i,u)]]$.

Can write down the corresponding Hamilton-Jacobi-Bellman equation and verification theorem.

$\Longrightarrow$ Optimal

$$u_i^*(t) \in \text{Argmax}\left(\sum_{j=1}^{n-m} p(j|i,\cdot)x_j^*(t) + x^* \sum_{j=n-m+1}^{n} p(j|i,\cdot)\right)$$

for $x < x^*$, and,

$$u_i^*(t) \in \text{Argmin}\left(\sum_{j=1}^{n-m} p(j|i,\cdot)x_j^*(t) + x^* \sum_{j=n-m+1}^{n} p(j|i,\cdot)\right)$$

for $x > x^*$.

($\Longrightarrow$ greatest 'push' towards x^*.)

References

1. VB, R. Makhijani, R. Sundaresan, **Asynchronous gossip for averaging and spectral ranking**, *IEEE J. Selected Topics in Signal Processing* 8(4), 2014.

2. VB, A. Karnik, U. Jayakrishnan Nair, S. Nalli, **Manufacturing consent**, to appear in *IEEE Transactions on Automatic Control*.

3. A. S. Mathkar, S. Phade, VB, **Nonlinear gossip**, in preparation.

I get by with a little help from undergrads/interns.

(V. Borkar)

$\cdots\cdots$ and I say, it's all right.

(George Harrison)

It's all over now, baby blue.

(Bob Dylan)

And I do appreciate you being around.

(Lennon-McCartney)