Naive Learning in Social Networks and the Wisdom of Crowds

by Benjamin Golub and Matthew O. Jackson

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Jakub Pawelczak

Paris 27.03.2018

Table of contents

Table of contents

PANTHÉON SORBONNE — UNIVERSITÉ PARIS 1

Social networks capture transfer of information, influence and opinions forming based on initial beliefs updated by imposed rule.

Questions

It is important to understand how beliefs and behaviors evolve over time, how they depend on the network structure and whether or not the outcome is efficient.

In paper written by Golub and Jackson they examined which social network structures society of agents, who communicate and update naively come to aggregate decentralized information correctly. They start with agents which receive independent noisy signal about true value and then communicate in network by taking average of neighbours. They call societies in which such convergence takes place *wise*.

Main results

- They characterize wise networks.
- They show that opinions converge to true state of nature if and only if most influential agent vanishes with grow of society.
- So wisdom can fail iff there is an agent whose degree number is nonvanishing.
- More generaly they proved that having bounded number of agents who are *prominent* (with nonvanishing attention from rest of agents in network) causes learning to fail.
- Main results At the end they gave examples that wisedom and speed of convergence are not related.

Table of contents

Definition of model

A finite set $N = (1, \dots, n)$ of agents interact according to social network with given adjacency matrix $g = [g_{ij}]$. Interaction patterns are capture by stochastic matrix $T = [T_{ij}]$ where agent i hears the opinion of the agents with whom he interacts and assigns precision g_{ij} to agent j. So from this kind of Bayesian updating we get interaction matrix $T_{ij} = \frac{g_{ij}}{\sum_k g_{ik}}$.
Agents update beliefs by repeatedly taking average of their neighbors belifs with T_{ij} being weight to agent i place on beilef of agent j in forming belief for next period.
With given initial $p^{(0)}$ belief we obtain recursive updating rule:

$$p^{(t)} = Tp^{(t-1)} \tag{1}$$

Initial belief is given by true value μ with noisy signal e_i (zero mean, nonnegative variance rv). so $p_i^{(0)} = \mu + e_i$

Convergence

Definition

Walk, path, trail, cycle.
Matrix T is strongly connected (irreducible) if there is a path in T from any node to any other.
Matrix T is aperiodic if GCD of lengths of cycles is zero.

Theorem

A stochastic matrix T is convergent ($\lim T^t p$ for all p) $\iff$ aperiodic and irreducible. Then there is unique left eigenvetor s corresponding to eigenvalue 1, so $s = sT$.

Examples

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{aperiodic} \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad s = (2/5, 2/5, 1/5)$$

Undirected networks with equal weights

In case of model when each agent equally split attention among neighbors ($T_{ij} = \frac{g_{ij}}{\sum_k g_{ik}}$), so when we have symmetric (unrealistic assumption which produces restrictions) relations between agents. In this case stationary distribution is form as follows

$$s_i = \frac{d_i(g)}{\sum_j d_j(g)} \tag{2}$$

so the influence is directly proportional to degree.

Table of contents

Definition of wisdom

There is a true state of nature $\mu \in [0,1]$ being fixed nr and nr of agents is n. At time $t = 0$ each agent i noticed signal/opinion $p_i^{(0)}$ which was μ disturbed by some noise e_i. This is the only moment when we have perturbation in model, but all next opinions are rv where the limit of this sequance is given by $p_i^{\infty}(n)$.

Definition

The sequence of stoch matrix growing with nr of players T(n) is *wise* if

$$plim_{n\to\infty} \max_{i\leq n} |p_i^{\infty}(n) - \mu| = 0 \tag{3}$$

The full characterization of wise society is given by theorem.

Theorem

If T(n) a sequence of convergent stochastic matrices then it is wise $\iff$ (s(n)) (sequence of influence vectors (stationary)) are such that $s_1(n) \to 0$.

Wisdom -remarks

So the society is wise iff most important agent's influence diminish with increase of society. THis is very important notion, society will fail to converge to truth if leader has too much power on forming opinions of others. In case of our example
Corollary Let g(n) be sequence of adjacency matrices. g(n) is wise $\iff$

$$\max_{1 \leq i \leq n} \frac{d_i(g)}{\sum_j d_j(g)} \to^n 0 \tag{4}$$

Table of contents

Definition

The group B is prominent in t steps (related to T) if

$$\forall i \notin B \qquad \sum_{j \in B} T_{ij} > 0 \tag{5}$$

So a group that is prominent in t steps is one such that each agent outside of it is influenced by at least someone in B group in t steps. So all people outside B pay attention to someone in group B.
The uniformly prominent family B_n is such that there exists threshold > 0 that min of sum in definition is $\geq \alpha$
Family B_n is finite if there exists $q > 0$ $\sup_n |B_n| \leq q$.

Prominent groups

Following theorem holds which gives negative condition on wisdom for prominent groups.

Theorem

If there is a finite, uniformly prominent family with respect to $T(n)$ then the sequence of networks is not wise

It's not full characterization, they gave example when wisdom fails and we do not have prominent group.By giving extra definition and condition on balancing (changing q by increasing to ∞ function of n) and minimal out-dispersion (which is for me superficial and without proper proof), they provide sufficient condition for wisdom:

Theorem

If T(n) is sequence of convergent stochastic matricies satisfying balance and minimal out-dispersion then it's wise

Table of contents

Speed of convergence

In this section they do not provide any bound for speed of convergence which is one of main drawbacks of paper. Especially lower bound would be interesting in this case. So I did my research and prove on my own:

Definition

Distance from stationty distribution in time t is defined as:

$$d(t) = \max_{x \in \Omega} ||P^t(x, \cdot) - \pi(\cdot)||_{tv} \tag{6}$$

Definition

Mixing time with respect to ϵ is minimal time after which we will be at least ϵ close to stationary:

$$t_{mix}(\epsilon) \min\{t \geq 0 : d(t) \leq \epsilon\} \tag{7}$$

My remarks

Theorem

For $\pi_{min} = \min_{x \in \Omega} \pi(x)$ mixing time is bounded by:

$$\left(\frac{1}{\gamma_*} - 1\right) log\left(\frac{1}{2\epsilon}\right) \overset{L}{\leq} t_{mix} \overset{P}{\leq} log\left(\frac{1}{\epsilon \pi_{min}}\right) \frac{1}{\gamma_*} \quad (8)$$

Examples

Lemma If $< f, 1 >_\pi = 0$ then

$$||Pf||_\pi \leq |\lambda| ||f||_\pi$$

Let f be eigenvector for $\lambda \neq 1$.Then $Pf = \lambda f$. eigen vectors are π-orthogonal so we use lemma:

$$0 =< f, 1 >_\pi = \sum \pi(y) f(y)$$

$$
\begin{aligned}
|\lambda^t f(x)| =& |P^t f(x)| = |\sum_{y\in\Omega} (P^t(x,y)f(y) - \pi(y)f(y))| \\
\leq& \sum_{y\in\Omega} |P^t(x,y) - \pi(y)||f(y)| \leq ||f||_\infty \sum_{y\in\Omega} |P^t(x,y) - \pi(y)| \\
=& ||f||_\infty \cdot 2 \cdot ||P^t(x,\cdot) - \pi(\cdot)||_{tv} \\
\leq& 2 \cdot ||f||_\infty \max_{x\in\Omega} ||P^t(x,\cdot) - \pi(\cdot)||_{tv} = 2 \cdot ||f||_\infty d(t)
\end{aligned} \tag{9}
$$

by taking sup with respect to x we obtain:

$$|\lambda^t| \leq 2d(t)$$

let's take $t = t_{mix}(\epsilon)$ so $d(t) \leq \epsilon$ hence:

$$|\lambda|^{t_{mix}} \leq 2\epsilon$$

$$\log(\frac{1}{2\epsilon}) \leq \log(\frac{1}{|\lambda|}) \cdot t_{mix}(\epsilon) \leq t_{mix}(\epsilon)(\frac{1}{|\lambda|} - 1)$$

To sum up we end up with:

$$t_{mix}(\epsilon) \geq \log(\frac{1}{2\epsilon})\frac{|\lambda|}{1-|\lambda|} \geq (\frac{1}{\gamma_*} - 1)\log(\frac{1}{2\epsilon})$$

To prove righthandside inequality we need following lemma.
Lemma2 : $\sum_{i=2}^{k} f_i(x)^2 \leq \frac{1}{\pi(x)}$

$$\begin{aligned} |\frac{P^t(x,y)}{\pi(y)} - 1| &= \sum_{i=2}^{k} |f_i(x) f_j(y) \lambda_i^t| \leq \sum_{i=2}^{k} |f_i(x) f_j(y)| \lambda_*^t \\ &\leq \lambda_*^t [\sum_{i=2}^{k} fi^2(x) \sum_{i=2}^{k} fi^2(y)]^{1/2} \leq \frac{\lambda^t}{\sqrt{\pi(x)\pi(y)}} \\ &\leq \frac{\lambda^t}{\pi_{min}} = \frac{(1-\gamma_*)^t}{\pi_{min}} \leq \frac{e^{-\gamma_* t}}{\pi_{min}} \end{aligned} \quad (10)$$

My remarks 4

$$d(t) \leq \max_{x\in\Omega} s_x(t) \leq \max_{x\in\Omega}\max_{y\in\Omega} |\frac{P^t(x,y)}{\pi(y)} - 1| = \frac{e^{-\gamma_* t}}{\pi_{min}} \tag{11}$$

$$d(t)\pi_{\min} \leq \frac{e^{-\gamma_* t_{mix}}}{\pi_{min}} \tag{12}$$

so:

$$\log(\frac{1}{\epsilon\pi_{\min}}) \leq \gamma_* t_{mix} \tag{13}$$

We proved theorem

Theorem

For $\pi_{min} = \min_{x\in\Omega} \pi(x)$ mixing time is bounded by:

$$\left(\frac{1}{\gamma_*} - 1\right) log\left(\frac{1}{2\epsilon}\right) \overset{L}{\leq} t_{mix} \overset{P}{\leq} log\left(\frac{1}{\epsilon\pi_{min}}\right)\frac{1}{\gamma_*} \tag{14}$$

Wisdom vs speed

The lack of any necessary relationship between convergence and wisdom can be seen via examples given in paper.

- First, consider the case where all agents weight each other equally. This society is wise and has immediate convergence.
- Second, consider a society where all agents weight just one agent. Here, we have immediate convergence but no wisdom
- Third, consider a setting where all agents place$1 - \epsilon$ weight on themselves and distribute the rest equally. This society is wise but can have arbitrarily slow convergence if ϵ is small enough.
- suppose all agents place $1 - \epsilon$ weight on themselves and the rest on one particular agent. Then there is neither wisdom nor fast convergence.

Table of contents

Final comments

- The main topic of this paper concerns what types of societies, whose agents get noisy signals at the beginning of the true value of some variable, are able to aggregate information in an approximately efficient way despite their naive and simple and decentralized updating. The existence of prominent groups or lead can mislead agents.
- The reason is simple opinion of prominent agents are overweighted and will dominate groups opinions over time. Even though authors gave sufficient conditions for wise networks definitions used in theorem are highly impractical and not useful.

Final comments 2

In authors opinion this paper gives more informative conditions on wisdom. But long line of previous papers (f.e. with observation learning , naive learning) suggests that sufficient conditions are hopelessly strong.Well here we pay the price.
In contrary to Acemoglu (2008) paper it has few similarities and differences driven by differences in agents' rationality.
Above all they forgot to:

- impose technical assumption of non negativity of initial belief vector
- form inequalities for speed of convergence
- try to simplify strong and higly impractical (in my opinion) condition on wise networks

Thank your for your attention! :)