WELCOME

Gossip Algorithms for Distributed Computing

Harish V

Presentation at the Weekly Group Meet,
Signal Processing for Communications Lab,
Indian Institute of Science,
29 September, 2012

Outline

- **Introduction**

Outline

- **Introduction**
- **An Overview on the Field of Interest**
  - The decentralized or distributed computation,
    a short historical perspective.
  - Applications
  - On the amount of literature
  - The elements of gossip algorithms
  - General analysis of gossip algorithms
  - The issues related to wireless
  - Few example works
  - The variant of consensus

Outline

- **Introduction**
- **An Overview on the Field of Interest**
  - The decentralized or distributed computation,
    a short historical perspective.
  - Applications
  - On the amount of literature
  - The elements of gossip algorithms
  - General analysis of gossip algorithms
  - The issues related to wireless
  - Few example works
  - The variant of consensus
- **The current work from our lab**
  - A consensus problem formulation
  - A solution and its characteristics

Outline

- **Introduction**
- **An Overview on the Field of Interest**
  - The decentralized or distributed computation,
    a short historical perspective.
  - Applications
  - On the amount of literature
  - The elements of gossip algorithms
  - General analysis of gossip algorithms
  - The issues related to wireless
  - Few example works
  - The variant of consensus
- **The current work from our lab**
  - A consensus problem formulation
  - A solution and its characteristics
- **Conclusions**

Before we begin...

Before we begin...

- We already saw a flavor of it last week..
  (ppt by Bharath, on work by D.Shah etal, JSTSP'11.)

Before we begin...

- We already saw a flavor of it last week..
  (ppt by Bharath, on work by D.Shah etal, JSTSP'11.)

- The work presents some of the recent developments in a broad class of problems.

Before we begin...

- We already saw a flavor of it last week..
  (ppt by Bharath, on work by D.Shah etal, JSTSP'11.)
- The work presents some of the recent developments in a broad class of problems.
- As we go along, we shall see where the work belongs to, in a much broader sense.

A Historical View

- The decentralized or distributed computation

A Historical View

- The decentralized or distributed computation – by wish or by order.

A Historical View

- The decentralized or distributed computation – by wish or by order.
- Dates back to more than 60 years old references..!

A Historical View

- The decentralized or distributed computation – by wish or by order.
- Dates back to more than 60 years old references..!
- The authoritative (& popularly cited) references of original work are
  - The team decision problem — Marschak etal. [1955]
  - The dissertation by Tsitsiklis at MIT [1984].
- An example: The hats problem (a well-known puzzle).

Few excerpts from Tsitsiklis dissertation, 1984

Questions

“ Given a set of *processors* with partial information, is it possible that they make satisfactory decisions (in certain sense) without communicating among them??”

“ If the given a set of *processors* with partial information must communicate, what is the least amount of communication required??”

Few excerpts from Tsitsiklis dissertation, 1984

Questions

" Given a set of *processors* with partial information, is it possible that they make satisfactory decisions (in certain sense) without communicating among them??"

" If the given a set of *processors* with partial information must communicate, what is the least amount of communication required??"

Results

" Optimal design of a decentralized system even in the absence of any dynamics, is very hard computationally."

" The basic, static problems of decentralized decision making are algorithmically hard (NP-hard) even though the corresponding centralized problem is *trivial*."

The techniques developed for his problem aimed for "*exchange of optimal decisions converging to consensus*".

- “Decentralized computation” by Tsitsiklis didn’t receive much attention when it was published [1984].

- "Decentralized computation" by Tsitsiklis didn't receive much attention when it was published [1984].
- In recent time around the year 2000, the topic received a considerable interest from several disciplines.

- "Decentralized computation" by Tsitsiklis didn't receive much attention when it was published [1984].
- In recent time around the year 2000, the topic received a considerable interest from several disciplines.
- "*Probably it is due to the fact that it has become hard to think of sensing etc problems in electronic systems, without involving a networked platform*" [intro to JSTSP,2011]

- "Decentralized computation" by Tsitsiklis didn't receive much attention when it was published [1984].
- In recent time around the year 2000, the topic received a considerable interest from several disciplines.
- "*Probably it is due to the fact that it has become hard to think of sensing etc problems in electronic systems, without involving a networked platform*" [intro to JSTSP,2011]
- "Gossip algorithms or distributed consensus algorithms" — current active variants of *decentralized computation*.

Applications

Span over a wide range of fields..

Applications

Span over a wide range of fields..

- **Wireless Networks:** Collaborative in-network processing (eg: spectrum-sensing in cognitive radio),
  - Reducing the bandwidth, energy, latency thereby increasing the lifetime of the system.
  - Reducing the redundancy and wastage of transmissions over centralized.
  - Surpassing the requirement of specialized routing service.

Applications

Span over a wide range of fields..

- **Wireless Networks:** Collaborative in-network processing (eg: spectrum-sensing in cognitive radio),
  - Reducing the bandwidth, energy, latency thereby increasing the lifetime of the system.
  - Reducing the redundancy and wastage of transmissions over centralized.
  - Surpassing the requirement of specialized routing service.
- **Computing:** Load balancing in distributed multi processors.

Applications

Span over a wide range of fields..

- **Wireless Networks:** Collaborative in-network processing (eg: spectrum-sensing in cognitive radio),
  - Reducing the bandwidth, energy, latency thereby increasing the lifetime of the system.
  - Reducing the redundancy and wastage of transmissions over centralized.
  - Surpassing the requirement of specialized routing service.
- **Computing:** Load balancing in distributed multi processors.
- **Automation and control:** Multi-agent coordination for vehicle formation, alignment, cooperative control.

A quantitative view on the literature

- "Distributed computation" has been a popular problem and there are over 10 thousand papers written on the subject.
- Variants of "Gossip algorithms" has seen a wide range of literature in the past decade with over 400 papers (80+ journals).

A quantitative view on the literature

- "Distributed computation" has been a popular problem and there are over 10 thousand papers written on the subject.
- Variants of "Gossip algorithms" has seen a wide range of literature in the past decade with over 400 papers (80+ journals).
- A special issue only on Gossip algorithms is issued by IEEE JSTSP in August,2011.

A quantitative view on the literature

- "Distributed computation" has been a popular problem and there are over 10 thousand papers written on the subject.
- Variants of "Gossip algorithms" has seen a wide range of literature in the past decade with over 400 papers (80+ journals).
- A special issue only on Gossip algorithms is issued by IEEE JSTSP in August,2011.
- The paper discussed in the last meet is one of the papers appeared in this issue.

A quantitative view on the literature

- "Distributed computation" has been a popular problem and there are over 10 thousand papers written on the subject.
- Variants of "Gossip algorithms" has seen a wide range of literature in the past decade with over 400 papers (80+ journals).
- A special issue only on Gossip algorithms is issued by IEEE JSTSP in August,2011.
- The paper discussed in the last meet is one of the papers appeared in this issue.
- The terms of "distributed computation", "Gossip", "Consensus" have become more generic.
- This presentation is an attempt to give an overall flavor..

Gossip Algorithm - Main elements

- Distributed computation of a function.

Gossip Algorithm - Main elements

- Distributed computation of a function.
- Distributed algorithms for nodes to update themselves to an optimal state, in order to optimize a central cost.

Gossip Algorithm - Main elements

- Distributed computation of a function.
- Distributed algorithms for nodes to update themselves to an optimal state, in order to optimize a central cost.
- **Major constraint**: Nodes can exchange the information (gossiping) only with their neighbors..

Gossip Algorithm - Main elements

- Distributed computation of a function.
- Distributed algorithms for nodes to update themselves to an optimal state, in order to optimize a central cost.
- **Major constraint**: Nodes can exchange the information (gossiping) only with their neighbors..
- "Even the simplest problems of decentralized decision making are *hard* enough".

Gossip Algorithm - Main elements

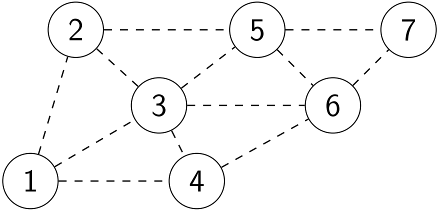
- Distributed computation of a function.
- Distributed algorithms for nodes to update themselves to an optimal state, in order to optimize a central cost.
- **Major constraint**: Nodes can exchange the information (gossiping) only with their neighbors..
- "Even the simplest problems of decentralized decision making are *hard* enough".
- We try to understand the inherent difficulties, by taking trivial problems in centralized setup to decentralized setup,. eg: Average, Product, Minimum, Maximum etc. (average is considered as a canonical choice)

Gossip Algorithm - Main elements

- Distributed computation of a function.
- Distributed algorithms for nodes to update themselves to an optimal state, in order to optimize a central cost.
- **Major constraint**: Nodes can exchange the information (gossiping) only with their neighbors..
- "Even the simplest problems of decentralized decision making are *hard* enough".
- We try to understand the inherent difficulties, by taking trivial problems in centralized setup to decentralized setup,. eg: Average, Product, Minimum, Maximum etc. (average is considered as a canonical choice)
- Readily extendable to a larger class of problems called *pair-wise* computable functions.

Contd.] Gossip Algorithms..

- "Who should communicate to whom and what, how often?"



Typical analysis for Algorithms:

- A proposal of communication protocol to converge to an optimal system state.
- Proof of convergence
- The convergence rate (mixing-time, second-eigenvalue..) computation
- Optimizing the convergence rate

An illustration of convergence:

- We consider,
  1. Average consensus
  2. A static network with N nodes
  3. At each time instant, an independently chosen subset of nodes exchange data and compute a linear average at each node.

An illustration of convergence:

- We consider,
  1. Average consensus
  2. A static network with N nodes
  3. At each time instant, an independently chosen subset of nodes exchange data and compute a linear average at each node.
- Can be illustrated as a matrix equation
  ($x(n)$ is a vector of data elements at all nodes at time n)

$$x(n+1) = W(n)x(n)$$

An illustration of convergence:

- We consider,
  1. Average consensus
  2. A static network with N nodes
  3. At each time instant, an independently chosen subset of nodes exchange data and compute a linear average at each node.
- Can be illustrated as a matrix equation ($x(n)$ is a vector of data elements at all nodes at time n)

$$x(n+1) = W(n)x(n)$$

- As only a partial information is conveyed throughout the network, it is required to perform an appropriate sequence of such communications (at least equal to the diameter of network).

An illustration of convergence:

- We consider,
  1. Average consensus
  2. A static network with N nodes
  3. At each time instant, an independently chosen subset of nodes exchange data and compute a linear average at each node.
- Can be illustrated as a matrix equation
  ($x(n)$ is a vector of data elements at all nodes at time n)

$$x(n+1) = W(n)x(n)$$

- As only a partial information is conveyed throughout the network, it is required to perform an appropriate sequence of such communications (at least equal to the diameter of network).
- For example, two of such operations are required if just three nodes are present in a line topology.

- If we consider after an arbitrary n consecutive operations,

$$x(n+1) = W(n).W(n-1)\ldots W(2).W(1).x(1) \qquad (1)$$
$$= \left(\prod_{i=1}^{n} W(i)\right) x(1) \qquad (2)$$

- When the appropriate randomization is employed, $W(n)$ become i.i.d. for each n. (example: randomized gossip, broadcast gossip etc.)

- As we are dealing with randomized algorithm, we are interested in the average performance of the system. Hence,

$$\mathbb{E}x(n+1) = \mathbb{E}\left(\prod_{i=1}^{n} W(i)\right)x(1) \tag{3}$$

$$= \left(\prod_{i=1}^{n} \mathbb{E}W(i)\right)x(1) \tag{4}$$

$$= (\mathbb{E}W)^n x(1) \tag{5}$$

- Hence the convergence of the system desires to have:

$$\lim_{n\to\infty} x(n) = \underline{1}x_{avg} \iff \lim_{n\to\infty} \mathbb{E}W^n = \frac{1}{N}\underline{1}.\underline{1}^T \tag{6}$$

THEOREM

If $P_{N\times N}$ is a doubly stochastic matrix,

$$\lim_{n\to\infty} P^n = \frac{1}{N}\underline{1}.\underline{1}^T \iff |\lambda_2(P)| < 1$$

Where $\lambda_i(A)$ denotes the i^{th} largest eigenvalue of the matrix A, in magnitude.

Proof:

Consider the difference,

$$P^n - \frac{1}{N}\underline{1}.\underline{1}^T = P^n - P\frac{1}{N}\underline{1}.\underline{1}^T, \quad \underline{1} \text{ is eigenvector} \tag{7}$$

$$= P^n - P^n\frac{1}{N}\underline{1}.\underline{1}^T \tag{8}$$

$$= P^n(I - \frac{1}{N}\underline{1}.\underline{1}^T) \tag{9}$$

$$= P^n(I - \frac{1}{N}\underline{1}.\underline{1}^T)^n, \quad \text{a projection matrix} \tag{10}$$

$$= (P - P\frac{1}{N}\underline{1}.\underline{1}^T)^n \tag{11}$$

$$= (P - \frac{1}{N}\underline{1}.\underline{1}^T)^n \tag{12}$$

Hence, LHS of theorem is true iff $|\lambda_1(P - \frac{1}{N}\underline{1}.\underline{1}^T)| < 1$.
Using simultaneous diagonalizability of P and $\frac{1}{N}\underline{1}.\underline{1}^T$,

$$\lambda_1(P - \frac{1}{N}\underline{1}.\underline{1}^T) = \lambda_2(P) \qquad \square$$

The immediate questions are

1. The speed of convergence [second eigenvalue(stoch. matx. convergence in its power) or mixing parameter(as a Markov chain)].
2. Optimize the speed of convergence for W.

Gossip Algorithms for Wireless:

Issues:

- Synchronized and asynchronized setups, while latter is more relevant.
- Time varying connectivity (time varying topology)
- Time varying values at nodes (advanced & is discussed by D.Shah etal., We saw it in last ppt)
- Noise & uncertainties in elements

Gossip Algorithms for Wireless:

Issues:

- Synchronized and asynchronized setups, while latter is more relevant.
- Time varying connectivity (time varying topology)
- Time varying values at nodes (advanced & is discussed by D.Shah etal., We saw it in last ppt)
- Noise & uncertainties in elements

Comments:

- A perfectly-synchronized setup with static-topology itself is complex in most cases.
- Like most wireless network problems, design of optimal protocols for fast convergence is an open problem.

Illustrating few examples

1) Boyd S., etal., "Randomized Gossip Algorithms", *IEEE Transactions on Information Theory*, June, 2006.

1. A gossip protocol — communication between a random pair of nodes
2. Synchronous model and asynchronous model. (Poisson timers at nodes)
3. Convergence Analysis
4. Rate of convergence by second eigenvalue
5. Optimizing the rate of convergence as an SDP

2) Aysal T. C., etal., “Broadcast Gossip Algorithms for Consensus”, *IEEE Transactions on Signal Processing*, July, 2009.

1. An asynchronous gossip protocol — A randomly selected node broadcasts its data to its neighbors. (Poisson timers at nodes)
2. Analysis of convergence in mean, mean-square sense.
3. Optimal mixing parameter computation.
4. Performance analysis — bound for mean square error and communication cost for achieving consensus.

3) Rajagopalan S., etal., "Distributed Averaging in Dynamic networks", *IEEE Journal of Selected Topics in Signal Processing*, August, 2011.

1. Seeking insights on effects in gossip based computations, when the network topology changes and the values themselves change
2. Bounded additive variations and Bounded multiplicative variations
3. Bharath has already discussed it in detail

The Problem of Consensus

- A special case of gossip algorithms, where the optimal state of each node is same.
- The version focused by most of the gossip algorithms at their initial level.

An interesting variant of consensus

- A higher layer problem so far.
- The physical layer was assumed to be well-serving, to address the problems related to higher layer precisely.
- All PHY issues such as communication noise and fading were ignored, given a link is present.

- Discrete value exchange between nodes: A natural question is to ask for consensus amidst noisy measurements.

- Discrete value exchange between nodes: A natural question is to ask for consensus amidst noisy measurements.
- There is a gap in the literature when each link is assumed to be noisy, faded.

- Discrete value exchange between nodes: A natural question is to ask for consensus amidst noisy measurements.
- There is a gap in the literature when each link is assumed to be noisy, faded.
- Some work already initiated in this direction. Its an enough challenging problem to start with.

- Discrete value exchange between nodes: A natural question is to ask for consensus amidst noisy measurements.
- There is a gap in the literature when each link is assumed to be noisy, faded.
- Some work already initiated in this direction. Its an enough challenging problem to start with.
- In the same lines of gossip algorithms, the question in its simplest form is posed asking nodes with discrete *binary* data values to achieve majority consensus among the nodes.

- Discrete value exchange between nodes: A natural question is to ask for consensus amidst noisy measurements.
- There is a gap in the literature when each link is assumed to be noisy, faded.
- Some work already initiated in this direction. Its an enough challenging problem to start with.
- In the same lines of gossip algorithms, the question in its simplest form is posed asking nodes with discrete *binary* data values to achieve majority consensus among the nodes.
- The problem of "Binary Consensus in Fading environment": my current focus. (A future presentation)

Thank You.