Diffusion Based EM Algorithm for Distributed Estimation of Gaussian Mixtures over Sensor Networks

WENG Yang

College of Mathematics, Sichuan University, China

DEC 2011

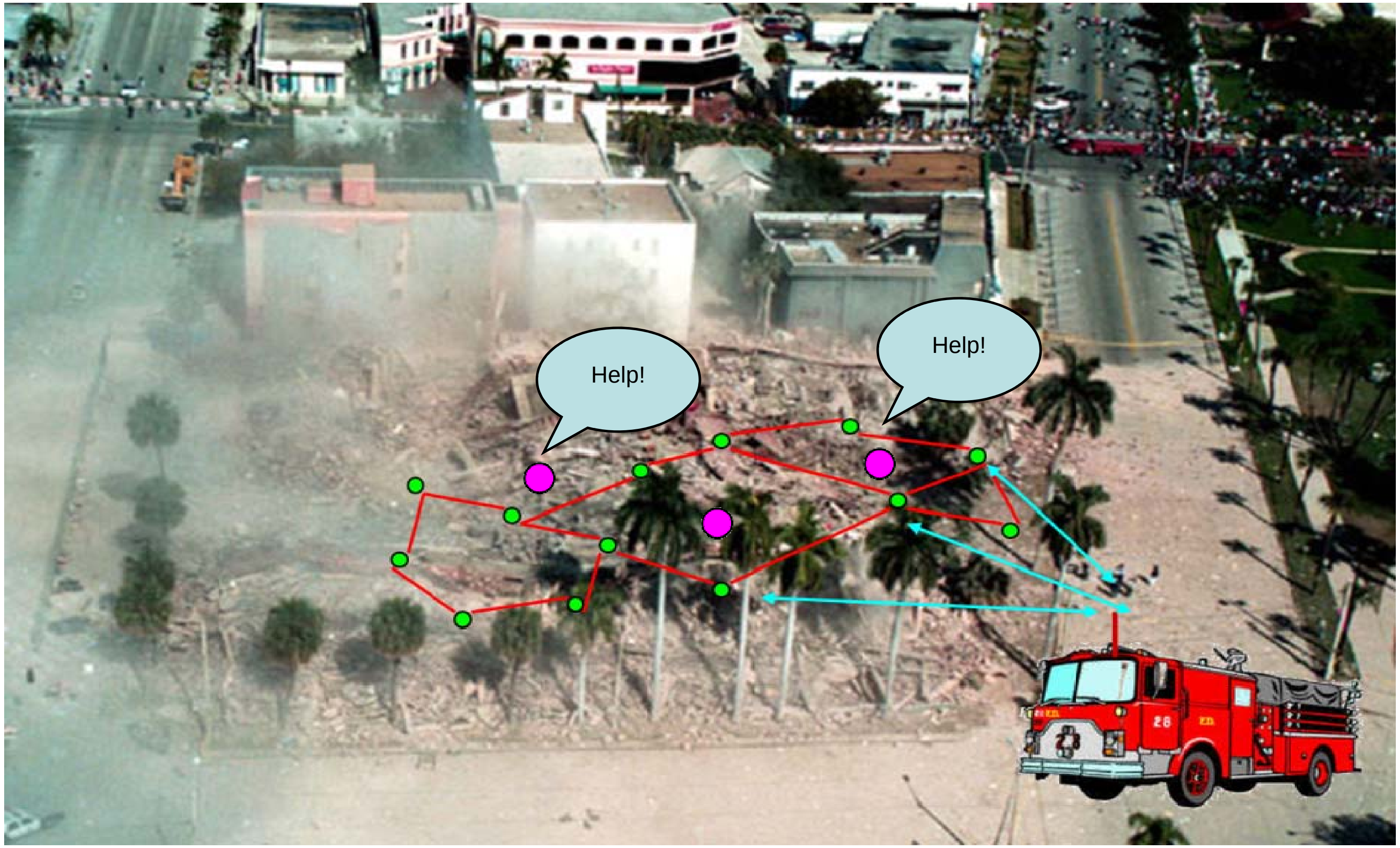
Outline

- Motivation
- Centralized EM algorithm for Gaussian mixture
- Distributed EM algorithm for Gaussian mixture
- Diffusion strategy for EM algorithm
- Stochastic approximation
- Simulations
- Conclusions and future work

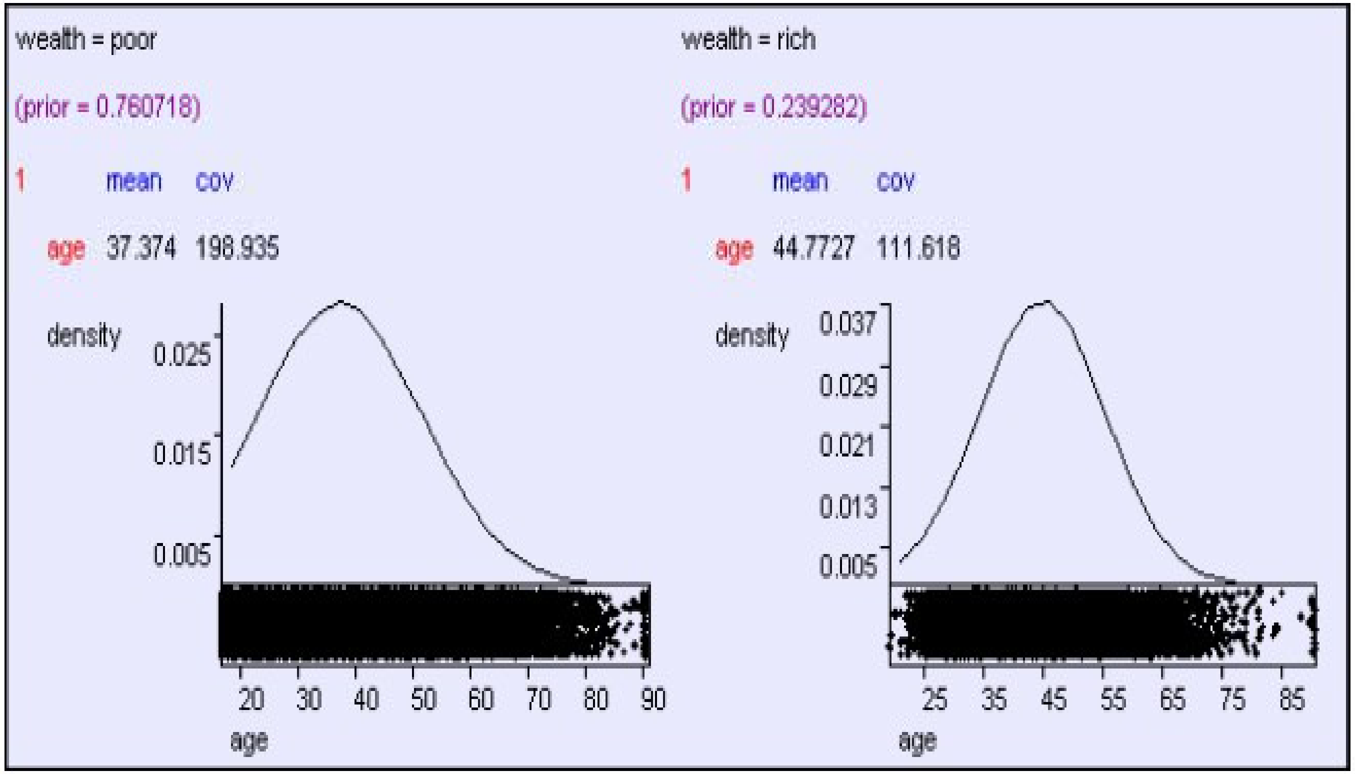
Outline

- Motivation
- Centralized EM algorithm for Gaussian mixture
- Distributed EM algorithm for Gaussian mixture
- Diffusion strategy for EM algorithm
- Stochastic approximation
- Simulations
- Conclusions and future work

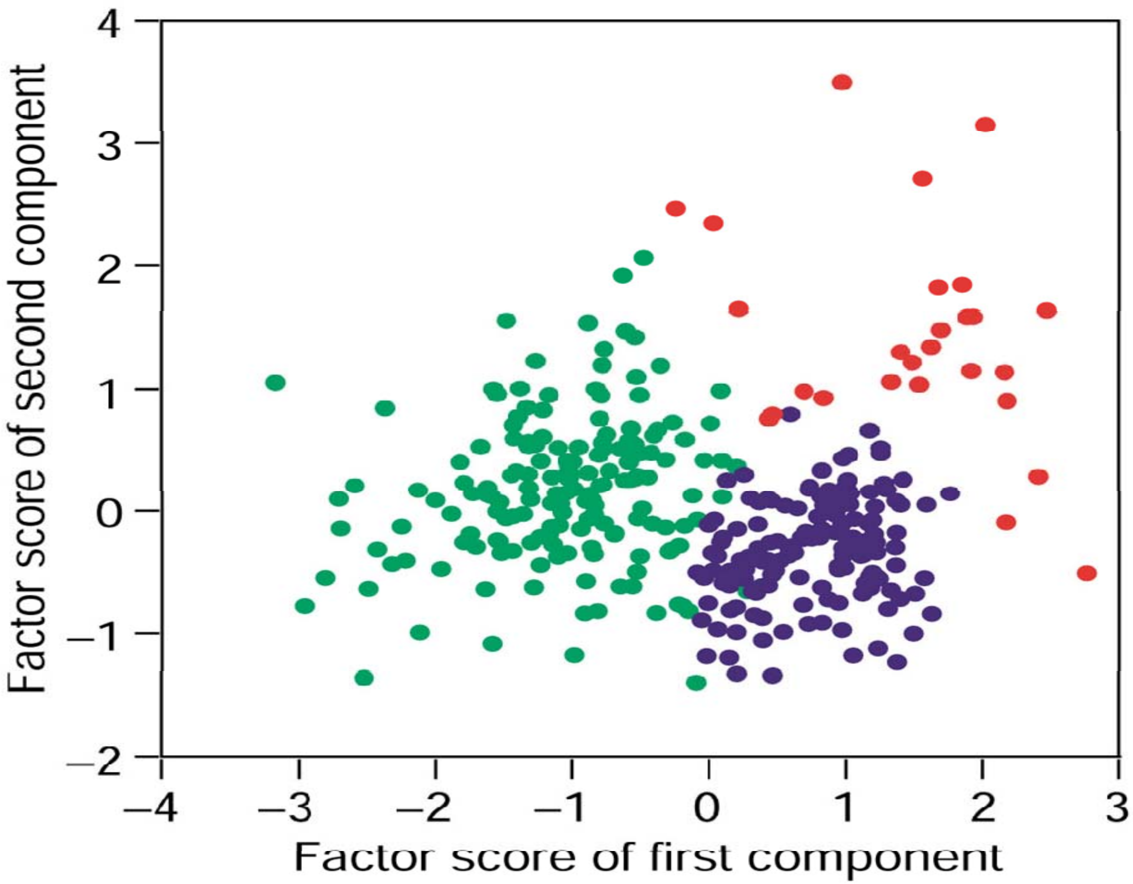
Multi-target localization in WSN



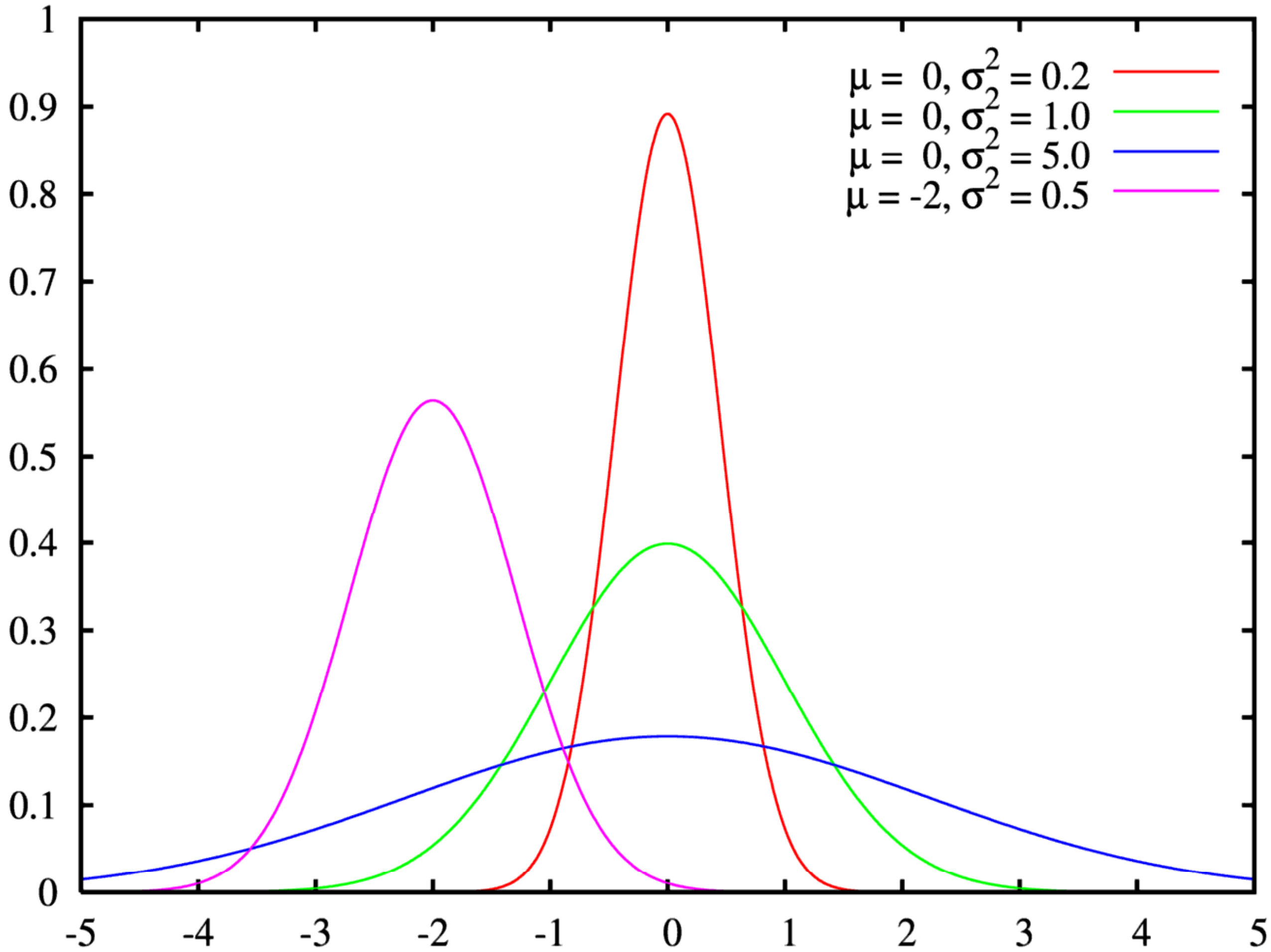
Predicting wealth from age



Parametric clustering of genes



Gaussian Mixture Model (GMM)



Motivations

- **Density estimation is central first step of data exploration in sensor network;**
- **Expectation-Maximization (EM) algorithm has been extensively exploited;**
- **The distributed processing method has many advantages in sensor network;**
- **The distributed manner of EM algorithm for Gaussian mixtures focus on the tradeoff between local processing and communication.**

Outline

- Motivation
- Centralized EM algorithm for Gaussian mixture
- Distributed EM algorithm for Gaussian mixture
- Diffusion strategy for EM algorithm
- Stochastic approximation
- Simulations
- Conclusions and future work

Centralized EM Algorithm for Gaussian Mixture Model in Sensor Network

- **A network with M sensor nodes;**
- **Each node has N_m observations, denoted as**

$$y_{mn}, m = 1, \cdots, M, n = 1, \cdots, N_m$$

- **observations are drawn form a Gaussian Mixtures with K components probabilities:** $\alpha_1, \cdots, \alpha_K$

$$y_{mn} \sim \sum_{j=1}^{K} \alpha_j \mathcal{N}(\mu_j, \Sigma_j)$$

Continued

- **Membership function**

$$\varepsilon^{t}_{mn,i} = p(z_{mn} = i \mid y_{mn}, \theta^{t})$$
$$= \frac{p(y_{mn} \mid z_{mn} = i, \theta^{t})\, p(z_{mn} = i \mid \theta^{t})}{\sum_{k=1}^{K} p(y_{mn} \mid z_{mn} = k, \theta^{t})\, p(z_{mn} = k \mid \theta^{t})}$$

- **E-step**

$$Q(\theta, \theta^{t}) = E_{z}\{\ln[\prod_{m=1}^{M}\prod_{n=1}^{N_m} p(y_{mn}, z \mid \theta)] \mid y_{mn}, \theta^{t}\}$$

- **M-step**

$$\theta^{t+1} = \arg\max_{\theta} Q(\theta, \theta^{t})$$

Continued

- The estimation of i-th component's parameters for the Gaussian mixture at step t+1

$$\mu_i^{t+1} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t} y_{mn}}{\sum_{m=1}^{M} \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t}}$$

$$\Sigma_i^{t+1} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t} (y_{mn} - \mu_i^{t+1})(y_{mn} - \mu_i^{t+1})^T}{\sum_{m=1}^{M} \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t}}$$

$$\alpha_i^{t+1} = \frac{1}{\sum_{m=1}^{M} N_m} \sum_{m=1}^{M} \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t}$$

Outline

- Motivation
- Centralized EM algorithm for Gaussian mixture
- Distributed EM algorithm for Gaussian mixture
- Diffusion strategy for EM algorithm
- Stochastic approximation
- Simulations
- Conclusions and future work

Distributed EM algorithm for Gaussian mixture

- **Incremental strategy** **(R. Nowak, TSP, 2003)**
  - A cyclic sequential communication path between sensor nodes is pre-determined;
  - The local statistics will be transmitted according to this path one sensor by one sensor.
- **Consensus strategy** **(D. Gu, TNN, 2008)**
  - A consensus filter will be implemented after the E-step at each iterative step, namely C-step (consensus step);
  - Each node will exchange local information with its neighbors until the agreement is reached.

Consensus strategy for parameter estimation

- The centralized EM algorithm enables the calculation of the global solution.
- Each node can implement the EM algorithm according to its own observations.
- How can we achieve the global solution for all nodes by transmitting the local information to the neighbors but not all observations?

Consensus strategy for mean estimation

$$\mu_i^{t+1} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t} y_{mn}}{\sum_{m=1}^{M} \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t}} = \frac{\frac{1}{M} \sum_{m=1}^{M} \beta_{m,i}^{t}}{\frac{1}{M} \sum_{m=1}^{M} \varepsilon_{m,i}^{t}}$$

- where: $\varepsilon_{m,i}^{t} = \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t}, \beta_{m,i}^{t} = \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t} y_{mn}$ are defined as Local Statistics.
- $\frac{1}{M} \sum_{m=1}^{M} \beta_{m,i}^{t}, \frac{1}{M} \sum_{m=1}^{M} \varepsilon_{m,i}^{t}$ can be obtained by a consensus filter.
- The variance and component probabilities can be obtained in the similar way.

Continued

$$\Sigma_i^{t+1} = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N_m}\varepsilon_{mn,i}^{t}(y_{mn}-\mu_i^{t+1})(y_{mn}-\mu_i^{t+1})^T}{\sum_{m=1}^{M}\sum_{n=1}^{N_m}\varepsilon_{mn,i}^{t}} = \frac{\frac{1}{M}\sum_{m=1}^{M}\gamma_{m,i}^{t}}{\frac{1}{M}\sum_{m=1}^{M}\varepsilon_{m,i}^{t}}$$

$$\alpha_i^{t+1} = \frac{1}{\sum_{m=1}^{M}N_m}\sum_{m=1}^{M}\sum_{n=1}^{N_m}\varepsilon_{mn,i}^{t} = \frac{M}{\sum_{m=1}^{M}N_m}\frac{1}{M}\sum_{m=1}^{M}\varepsilon_{m,i}^{t}$$

- where: $\gamma_{m,i}^{t} = \sum_{n=1}^{N_m}\varepsilon_{mn,i}^{t}(y_{mn}-\mu_i^{t})(y_{mn}-\mu_i^{t})^T$ is Local Statistics.

Disadvantages of consensus strategy

- **The convergence rates for consensus step will degrade as the number of sensor nodes increase.**
- **The consensus of global information for all the sensor nodes will be achieved with quite a bit of communications to the neighbors.**

Outline

- Motivation
- Centralized EM algorithm for Gaussian mixture
- Distributed EM algorithm for Gaussian mixture
- Diffusion strategy for EM algorithm
- Stochastic approximation
- Simulations
- Conclusions and future work

Diffusion strategy for parameter estimation

- **E-step:**

$$\varepsilon_{m,i}^{t} = \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t}, \beta_{m,i}^{t} = \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t} y_{mn}, \gamma_{m,i}^{t} = \sum_{n=1}^{N_m} \varepsilon_{mn,i}^{t} (y_{mn} - \mu_i^t)(y_{mn} - \mu_i^t)^T$$

- **D-step:**

$$\varepsilon_{m,i}^{t} = \frac{1}{|\mathcal{N}_i|} \sum_{l \in \mathcal{N}_m} \varepsilon_{l,i}^{t}, \beta_{m,i}^{t} = \frac{1}{|\mathcal{N}_i|} \sum_{l \in \mathcal{N}_m} \beta_{l,i}^{t}, \gamma_{m,i}^{t} = \frac{1}{|\mathcal{N}_i|} \sum_{l \in \mathcal{N}_m} \gamma_{l,i}^{t}$$

- **M-step:**

$$\mu_{m,i}^{t} = \frac{\beta_{m,i}^{t}}{\varepsilon_{m,i}^{t}}, \Sigma_{m,i}^{t} = \frac{\gamma_{m,i}^{t}}{\varepsilon_{m,i}^{t}}, \alpha_{m,i}^{t} = \frac{\varepsilon_{m,i}}{N_m}$$

Advantages of diffusion strategy

- **Each node only communicate the local statistics to neighboring nodes at each iterative step;**
- **Each node update the local statistics according to the information received from its neighbors;**
- **The consensus for local statistics doesn't need to be achieved which can save much communications.**

Outline

- Motivation
- Centralized EM algorithm for Gaussian mixture
- Distributed EM algorithm for Gaussian mixture
- Diffusion strategy for EM algorithm
- Stochastic approximation
- Simulations
- Conclusions and future work

Stochastic approximation

- **Diffusion distributed EM algorithm (DDEM) can be considered as Robbins-Monro stochastic approximation;**
- **Notation for weighted mean:**

$$\langle\langle f(y)\rangle\rangle_i(N) = \eta(N)\sum_{n=1}^{N}\left(\prod_{m=n+1}^{N}\lambda(s)\right) f(y_n)\, p(i \mid y_n, \theta^{n-1}),$$

$$\eta(N) = \left(\sum_{n=1}^{N}\prod_{m=n+1}^{N}\lambda(s)\right)^{-1}$$

Continued

- **At time step t, the estimation can be written as:**

$$\alpha_i^t = \langle\langle 1 \rangle\rangle_i (t), \mu_i^t = \frac{\langle\langle y \rangle\rangle_i (t)}{\langle\langle 1 \rangle\rangle_i (t)},$$

$$\Sigma_i^t = \frac{\langle\langle (y-\mu_i)(y-\mu_i) \rangle\rangle_i (t)}{\langle\langle 1 \rangle\rangle_i (t)}$$

- **Local statistics**

$$\phi_{m,i}^t \triangleq \{\varepsilon_{m,i}^t, \beta_{m,i}^t, \gamma_{m,i}^t\}$$

$$= \{\langle\langle 1 \rangle\rangle_i (t), \langle\langle y \rangle\rangle_i (t), \langle\langle (y-\mu_i)(y-\mu_i)^T \rangle\rangle_i (t)\}$$

Continued

- **On-line EM algorithm has the abstract form:**

$$\delta\phi_m^t = \phi_m^t - \phi_m^{t-1} = \eta(t)[F(y,\theta^{t-1}) - \phi_m^{t-1}]$$

$$\theta_m^t = H(\phi_m^t)$$

- **First order dynamics:** $\dot{z}(t) = -Lz(t)$
- **DDEM can be formulated by the first order dynamics in the discrete form:**

$$\delta\phi_m^t = \phi_m^t - \phi_m^{t-1}$$

$$= \eta(t)[F(y,\theta^{t-1}) - \phi_m^{t-1} + \sum_{n\in\mathcal{N}_m} (\phi_n^{t-1} - \phi_m^{t-1})]$$

Continued

- **Diffusion on-line EM algorithm can be written as:**

$$\delta\phi_m^t = \eta(t)\left(E[F(y,H(\phi_m^{t-1}))]_\rho - \phi_m^{t-1}\right) + \eta(t)\zeta(y_t,\overline{\phi}^{t-1})$$

- **where:** $\phi_m = E[F(y,\theta)]_\rho, \theta_m = H(\phi_m),$

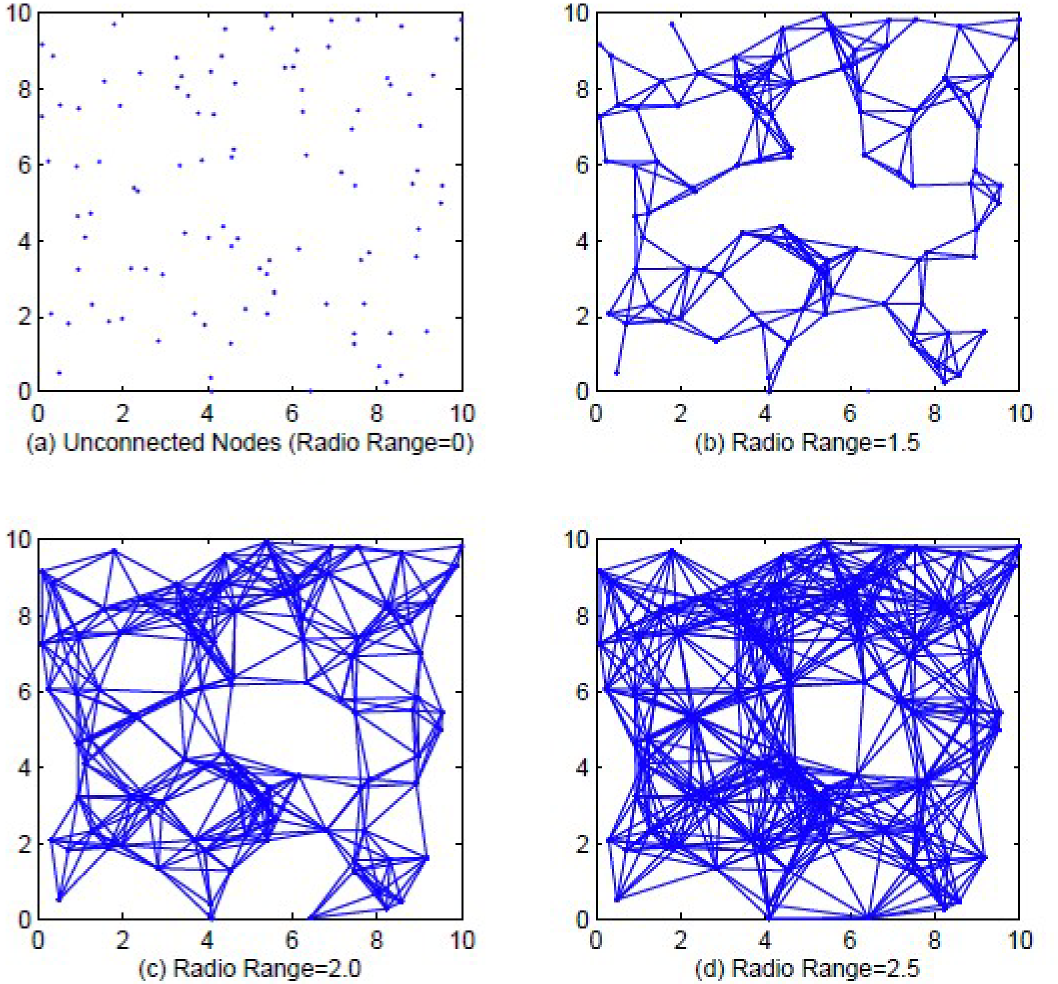
$$\zeta(y_t,\overline{\phi}^{t-1}) = F(y_t,H(\phi_m^{t-1})) -$$
$$E[F(y,H(\phi_m^{t-1}))]_\rho + \sum_{l\in\mathcal{N}_m^{t-1}} (\phi_l^{t-1} - \phi_m^{t-1})$$

- **The above formula has the same form as the Robbins-Monro stochastic approximation**

Outline

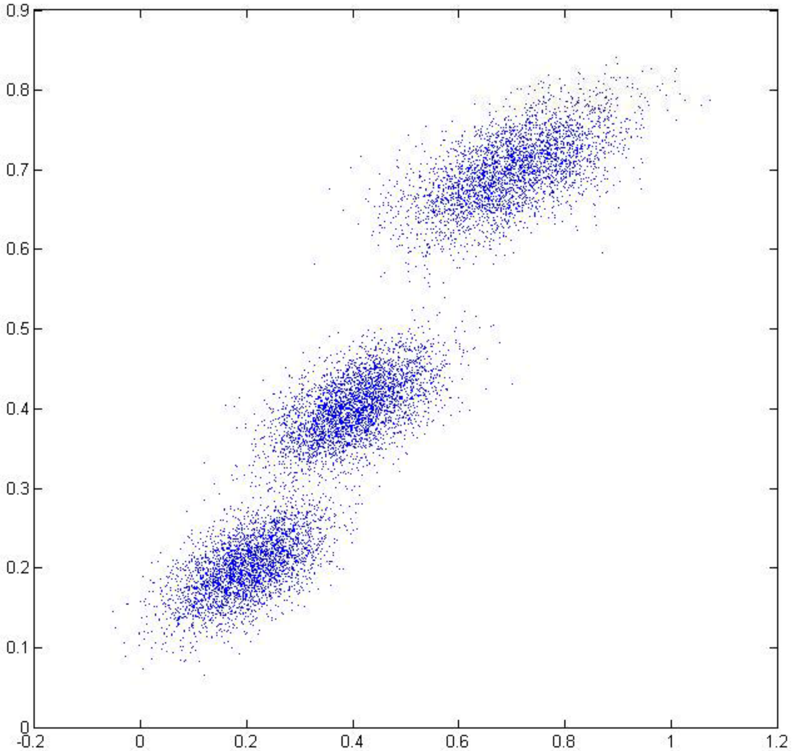
- Motivation
- Centralized EM algorithm for Gaussian mixture
- Distributed EM algorithm for Gaussian mixture
- Diffusion strategy for EM algorithm
- Stochastic approximation
- Simulations
- Conclusions and future work

Nodes Topology

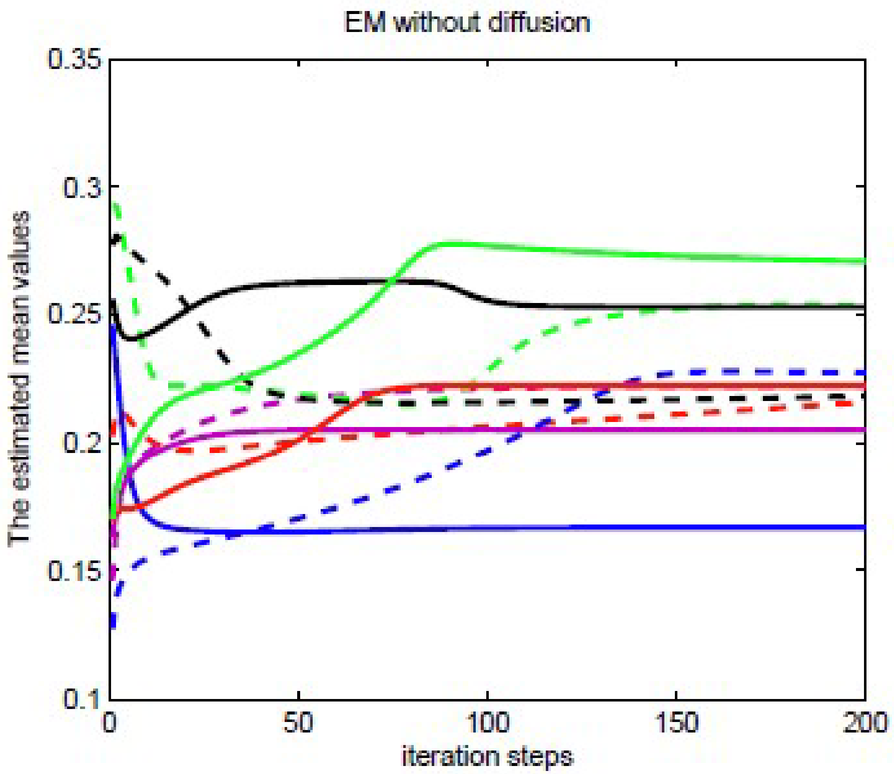


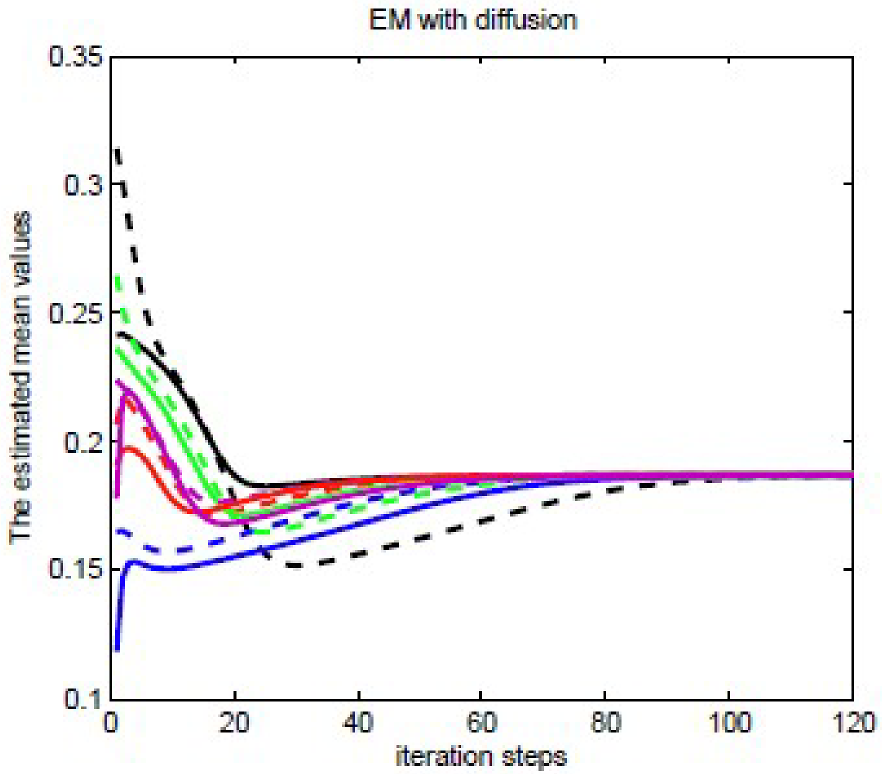
Data distribution

- In the first 30 nodes,90% observations come from the first component, the rest 10% evenly from the other two components;
- In the next 40 nodes,80% observations come from the second component, the rest 20% evenly from the other two components;
- In the first 30 nodes,90% observations come from the third component, the rest 10% evenly from the other two components

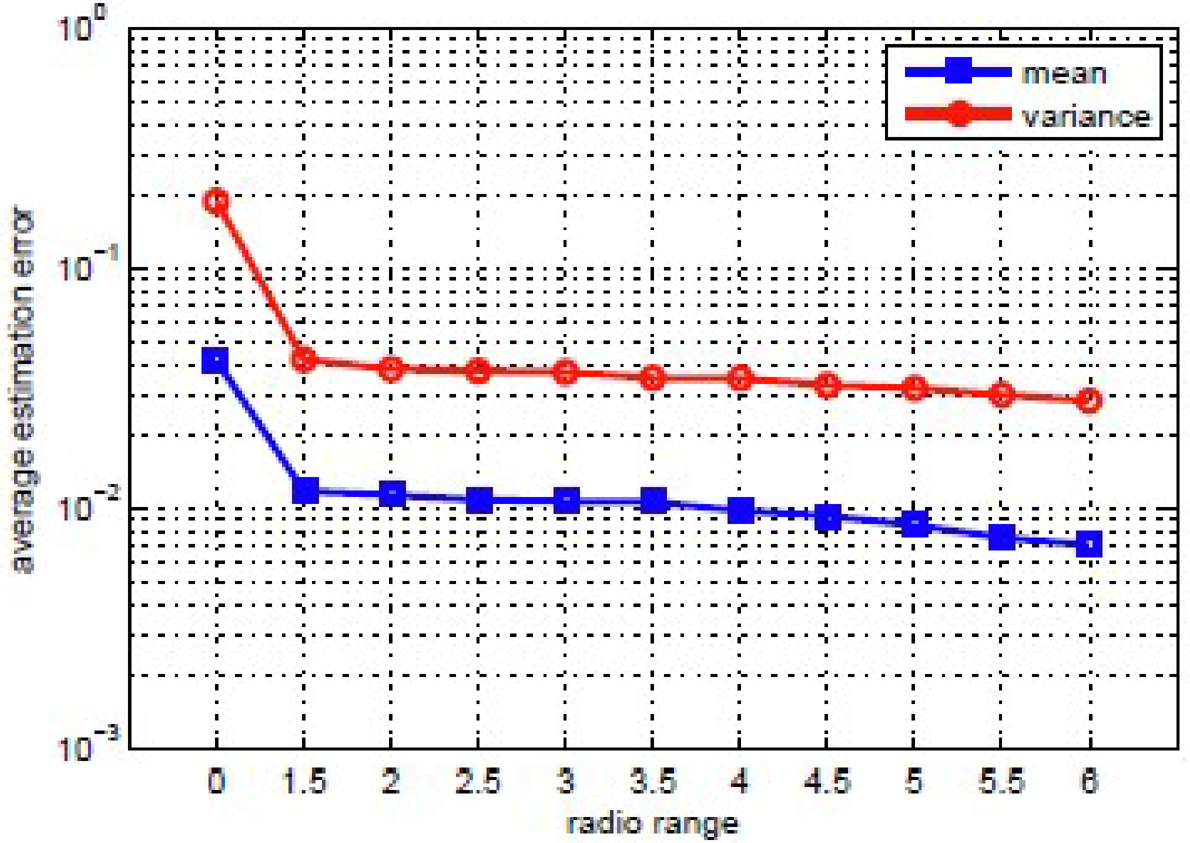


EM algorithm with and without diffusion

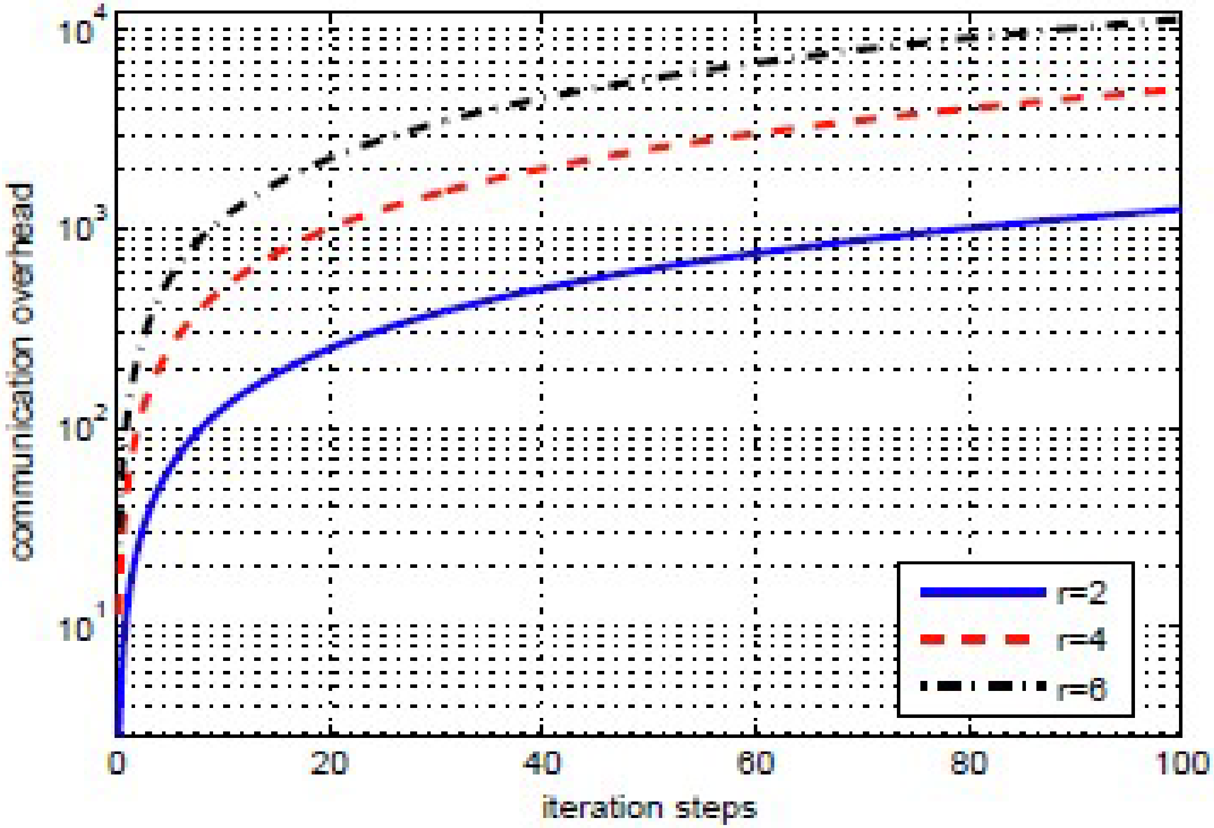




Estimation performance with different communication range



Communication overhead with different communication range



Outline

- **Motivation**
- **Centralized EM algorithm for Gaussian mixture**
- **Distributed EM algorithm for Gaussian mixture**
- **Diffusion strategy for EM algorithm**
- **Stochastic approximation**
- **Simulations**
- **Conclusions and future work**

Conclusions and future work

- **Each node only communicate the local statistics to neighboring nodes at each iterative step while the consensus strategy requires much more amount of communication to achieve consensus;**
- **Our method can be considered as the Robbins-Monro stochastic approximation to the maximum likelihood estimation for Gaussian Mixture;**
- **Is there exist some simple estimation method can substitute the ML estimation?**