# 网络上的分布式变分贝叶斯学习

华俊豪

统计信息与图像处理研究中心，信息与电子工程学院，
浙江大学

huajh@zju.edu.cn

# 研究方向



- 研究课题：网络上的**分布式变分贝叶斯**算法研究

关键词：贝叶斯学习，分布式信号处理，传感器网络

理论+应用

# 纲要

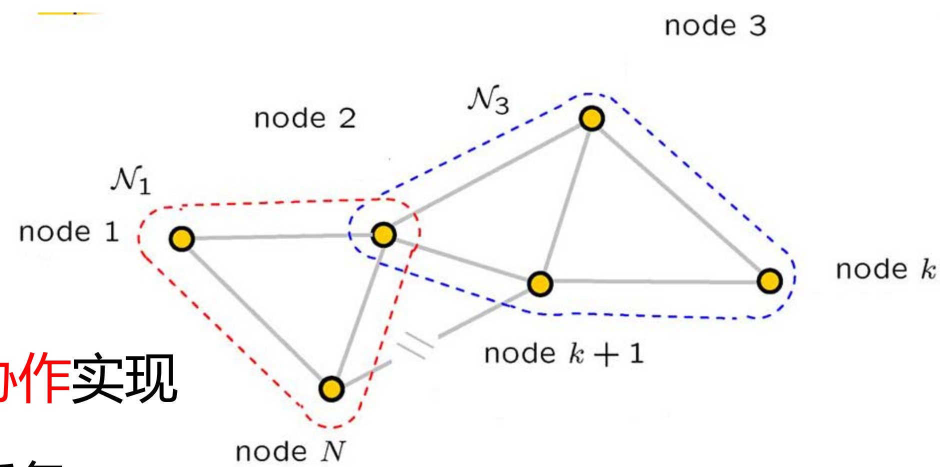
- 研究背景
- 研究现状与研究动机
- 分布式变分贝叶斯算法
- 算法应用

# 研究背景：网络化模型



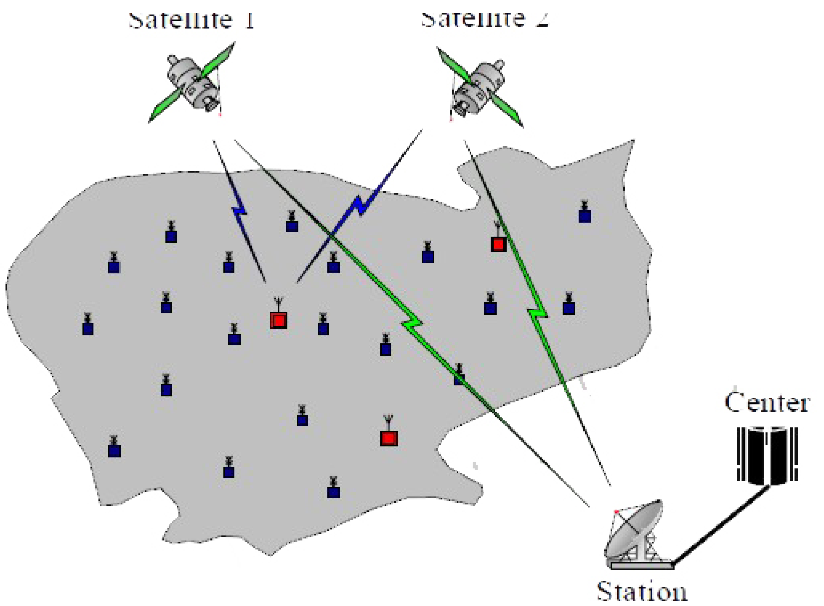
## ❑ 网络化模型

- ❑ 由分布在一定区域内的节点自组织成网络构成。
- ❑ 节点具有采集和处理数据的能力。
- ❑ 受能量，带宽和隐私保护的限制，各节点仅有，
  - 有限的计算能力，
  - 有限的通信能力，
  - 受保护的局部数据。
- ❑ **目的**：通过节点之间的相互协作实现分布式估计、推断、学习等任务。

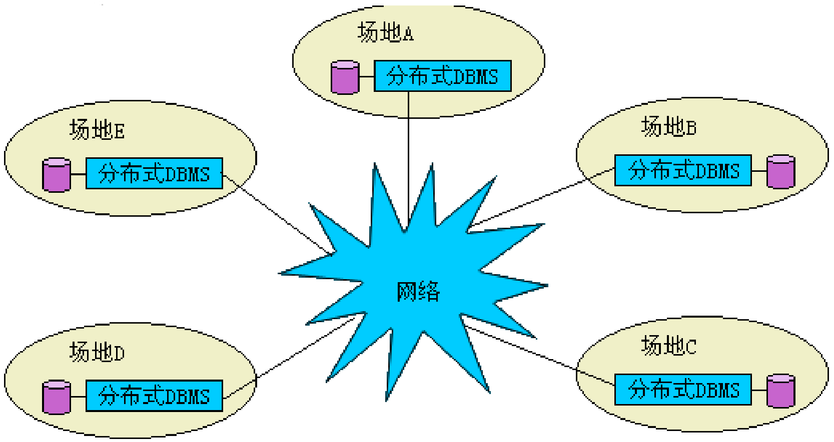


# 应用背景

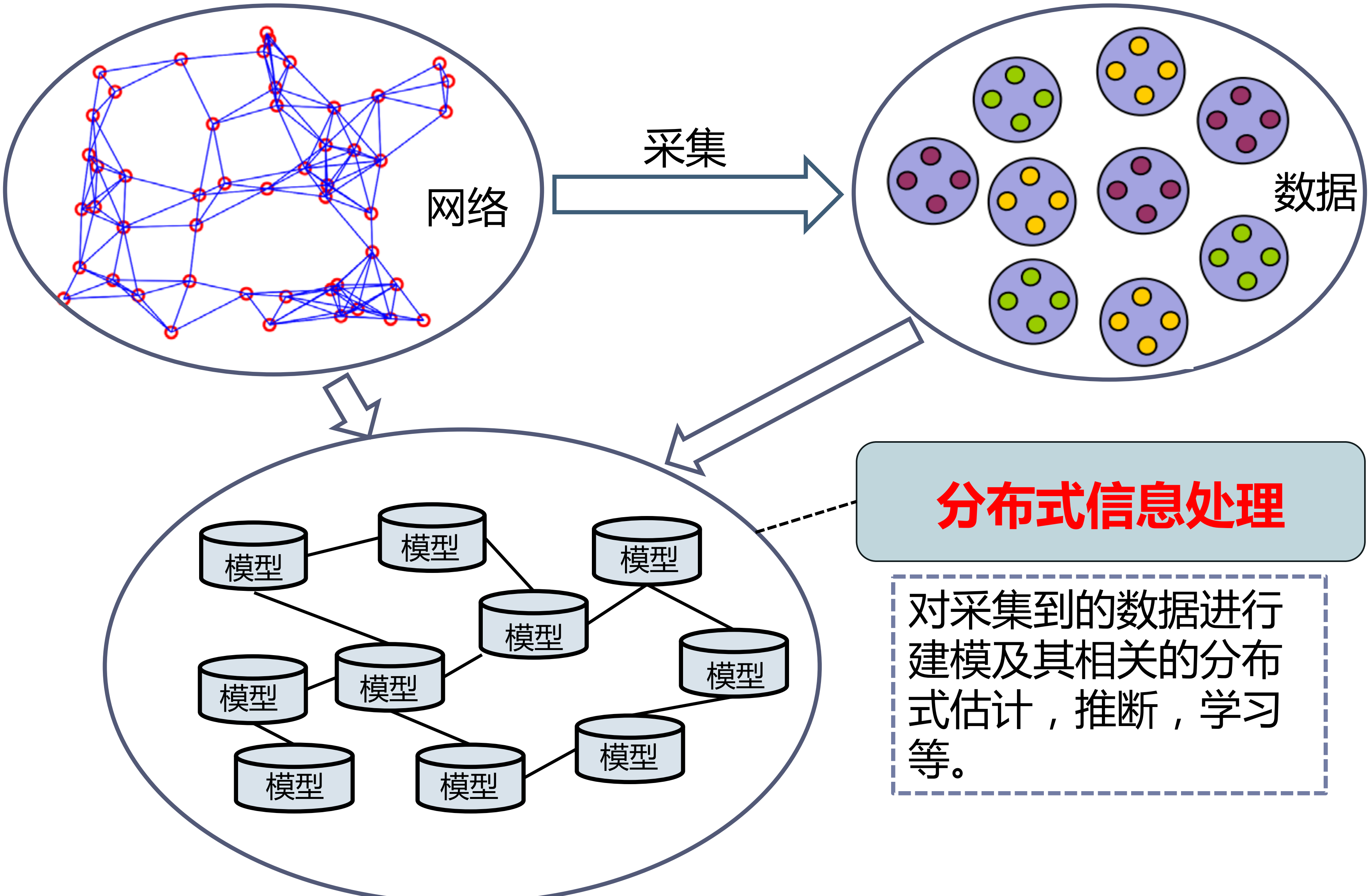
- （无线）传感器网络
  - 军事、农业、环境
  - 定位、跟踪、监测




- 分布式储存和计算机集群
  - 互联网公司、医院、政府机构
  - 海量数据挖掘、机器学习
  - 大数据



# 分布式信息处理




**分布式信息处理**

对采集到的数据进行建模及其相关的分布式估计，推断，学习等。

# 纲要



- ❑ 研究背景
- ❑ **研究现状与研究动机**
- ❑ 分布式变分贝叶斯算法
- ❑ 算法应用

# 分布式信息处理研究现状



- ❑ 涵盖各类学习问题
  - ❑ 分布式参数估计
  - ❑ 分布式状态跟踪
  - ❑ 分布式聚类/分类
  - ❑ 分布式检测等
- ❑ 统计模型的设计：大多是基于频率统计
  - ❑ 如统计假设检验，置信区间
  - ❑ 最小二乘估计，GMM聚类
  - ❑ 采用极大似然，EM算法

不考虑先验信息，强调数据出现频率或比例，从样本数据中得出结论

# 分布式信息处理研究现状

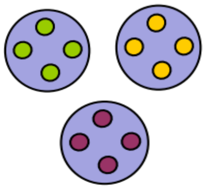
- ❑ 贝叶斯建模
  - ❑ 用一个先验分布来表达我们对未知数据的不确定性
  - ❑ 能提供一种优雅的途径对大规模数据及其潜在的结构进行分析。
  - ❑ 网络上的贝叶斯推断研究现状
    - 信息传递或置信传播(belief propagation)

      缺点：需重构网络结构，生成树算法

    - 非参数化置信传播 (Monte Carlo +particle filtering)

      缺点：计算量大

# 变分贝叶斯



**统计建模/推断**

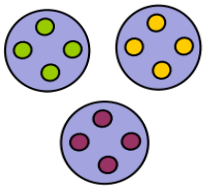
频率建模/推断

极大似然法

易过拟合，倾向复杂模型，可计算性差。

# 变分贝叶斯





统计建模/推断

频率建模/推断

极大似然法

易过拟合，倾向复杂模型，可计算性差。

贝叶斯建模/推断

可避免过拟合，能控制模型复杂性，但可能产生不可计算的积分。

近似技术

采样，Laplace等

**变分贝叶斯 (Variational Bayes)**

Thomas Bayes (1701-1761)

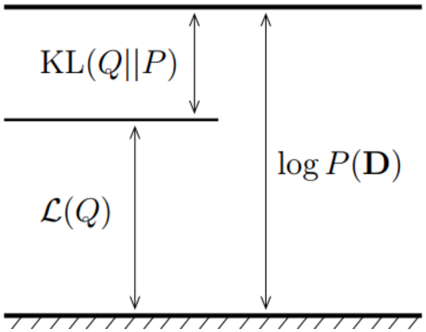
$$P(X|Y) = \frac{P(Y|X)P(X)}{\boxed{P(Y)}}$$

$$\boxed{P(Y) = \int P(Y|X)P(X)dX}$$

# 变分贝叶斯



- 原理
  - 用一类简单可计算（tractable)的分布Q来近似真实分布P。
  - 通过最小化KL散度，优化得到最优的变分分布Q。



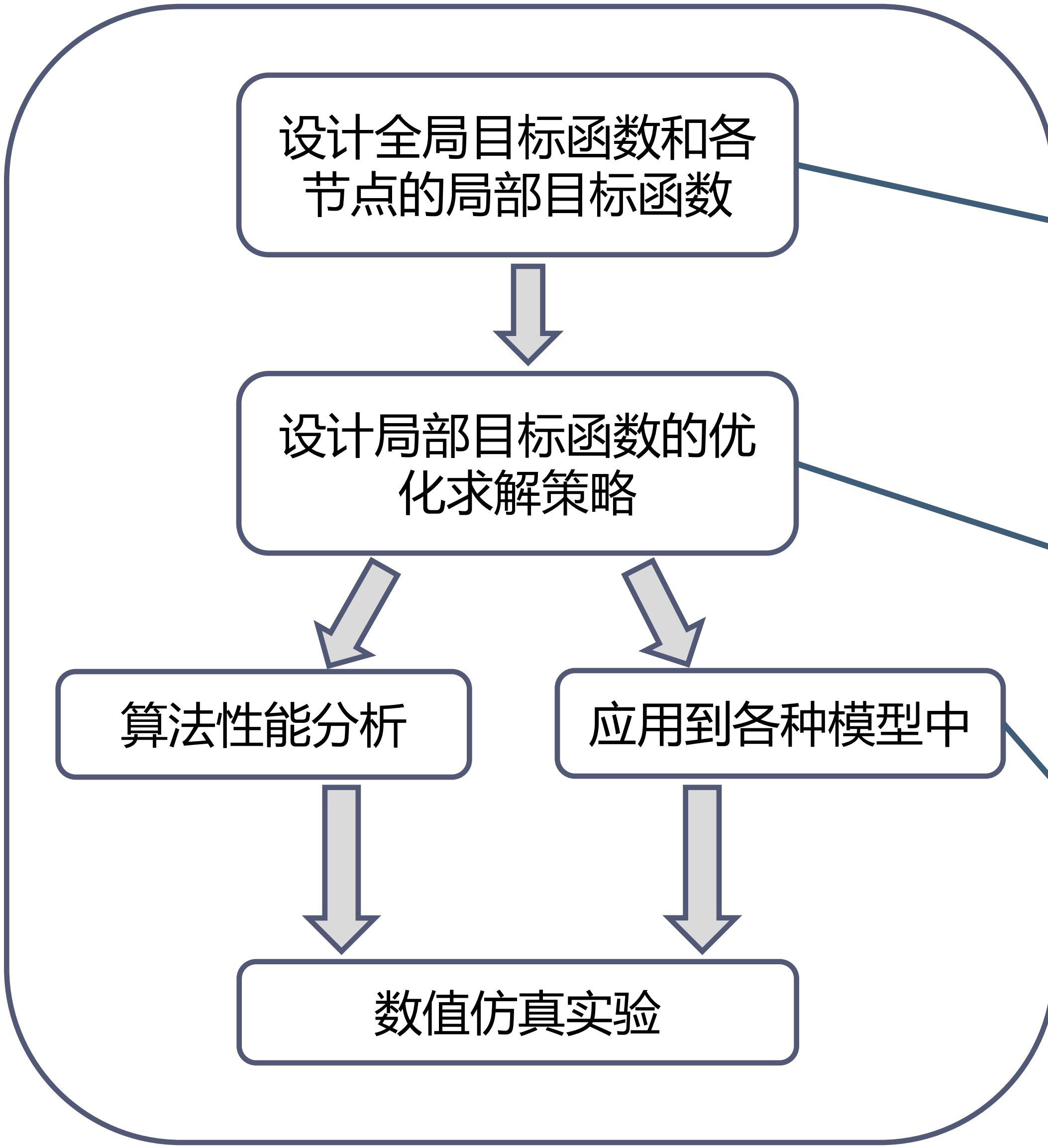

- 主要优势
  - 能得到参数和隐变量后验概率的解析表达式。
  - 能推导出边缘似然函数（证据）的下界，用于模型选择。
  - 同等精度下，变分贝叶斯方法比采样方法更快。
- 应用
  - 各类混合模型，概率主题模型，矩阵分解和时间序列分析等等。

# 纲要



- ❑ 研究背景
- ❑ 研究现状与研究动机
- ❑ **分布式变分贝叶斯算法**
- ❑ 算法应用

# 分布式变分贝叶斯算法设计流程




针对特定的分布族，选择合适的优化对象。

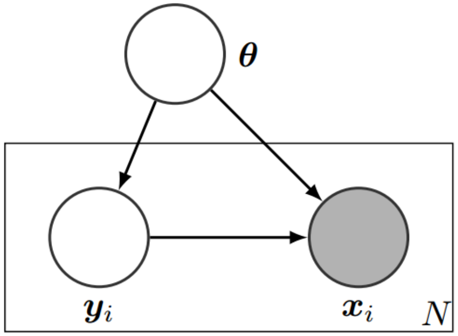
选择合适的优化方法和信息融合策略：

- 学习性能
- 无需传递原始数据
- 计算复杂度、通信复杂度

如高斯混合模型。

# 分布式变分贝叶斯框架

- 网络上的分布式变分贝叶斯框架
  - 将全局目标函数分解为各局部目标函数的平均




$$\begin{aligned}\mathcal{L}(Q(\boldsymbol{z})) &= \mathbb{E}_Q[\log P(\boldsymbol{z},\boldsymbol{x})] + \mathbb{H}[Q(\boldsymbol{z})]\\ &= \frac{1}{N}\sum_{i=1}^{N} N\mathbb{E}_Q[\log P(\boldsymbol{x}_i|\boldsymbol{z})] + \mathbb{E}_Q[\frac{P(\boldsymbol{z})}{Q(\boldsymbol{z})}]\\ &= \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_i(Q(\boldsymbol{z})),\end{aligned}$$

  - 考虑变分分布属于在共轭指数分布族。
  - 将优化变量从变分分布等价地换成自然参数向量。
  - 得到一般意义下的交替迭代框架：

$$\text{VBE}: \boldsymbol{\phi}_{y_i}^{*} = \arg\max_{\boldsymbol{\phi}_{y_i}} \mathcal{L}_i(\boldsymbol{\phi}_{y_i}, \boldsymbol{\phi}_{\theta}^{*}), \forall\, i = 1, \ldots, N, \quad (17a)$$

$$\text{VBM}: \boldsymbol{\phi}_{\theta}^{*} = \arg\max_{\boldsymbol{\phi}_{\theta}} \boxed{\sum_{i=1}^{N} \mathcal{L}_i(\boldsymbol{\phi}_{y_i}^{*}, \boldsymbol{\phi}_{\theta})}. \quad (17b)$$

分布式计算

# 两种分布式算法

- 分布式随机变分贝叶斯算法(dSVB)
  - VBM求解方案：基于梯度的随机变分推断 + 扩散策略

$$
\begin{aligned}
\varphi_{\theta,i}^{t} &= \boldsymbol{\phi}_{\theta,i}^{t-1} + \eta_t\left(\boldsymbol{\phi}_{\theta,i}^{*,t} - \boldsymbol{\phi}_{\theta,i}^{t-1}\right), \\
\boldsymbol{\phi}_{\theta,i}^{t} &= \sum_{j\in\mathcal{N}_i\cup\{i\}} w_{ij}\varphi_{\theta,j}^{t}.
\end{aligned}
$$

一般性：
共轭指数分布族

- 基于ADMM的分布式变分贝叶斯算法(dVB-ADMM)
  - VBM求解方案：定义一个等价的优化问题 + ADMM优化

$$
\begin{aligned}
\min_{\boldsymbol{\phi}_{\theta,i},\boldsymbol{\varphi}_{\theta,i,j}} \quad & \frac{1}{2}\sum_{i=1}^{N}\|\boldsymbol{\phi}_{\theta,i} - \boldsymbol{\phi}_{\theta,i}^{*}\|_F^2 \\
\text{s.t.} \quad & \boldsymbol{\phi}_{\theta,i} = \boldsymbol{\varphi}_{\theta,i,j}, i = 1,\ldots,N, j\in\mathcal{N}_i, \\
\text{s.t.} \quad & \boldsymbol{\varphi}_{\theta,i,j} = \boldsymbol{\phi}_{\theta,j}, i = 1,\ldots,N, j\in\mathcal{N}_i.
\end{aligned}
$$

# 纲要



- 研究背景
- 研究现状与研究动机
- 分布式变分贝叶斯算法
- **算法应用**

# 高斯混合模型



- 高斯混合模型: $p(\boldsymbol{x}_{ij}|\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Lambda}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}_{ij}|\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k^{-1}),$
- 应用于分布式**密度估计**问题，**聚类**问题
- 多层贝叶斯模型

$$P(\{\boldsymbol{x}_i\}_N|\boldsymbol{y}_i,\boldsymbol{\mu},\boldsymbol{\Lambda}) = \prod_{j=1}^{N_i}\prod_{k=1}^{K} \mathcal{N}(\boldsymbol{x}_{ij}|\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k^{-1})^{N\cdot y_{ijk}},$$

$$P(\boldsymbol{y}_i|\boldsymbol{\pi}) = \prod_{j=1}^{N_i} \mathrm{Mult}(1,\boldsymbol{\pi}),$$

$$P(\boldsymbol{\pi}) = \mathrm{Dir}(K,\alpha_0),$$

$$P(\boldsymbol{\mu}|\boldsymbol{\Lambda}) = \prod_{k=1}^{K} \mathcal{N}(\boldsymbol{\mu}_0,(\beta_0\boldsymbol{\Lambda}_k)^{-1}),$$

$$P(\boldsymbol{\Lambda}) = \prod_{k=1}^{K} \mathcal{W}(\boldsymbol{W}_0,\nu_0).$$

- 变分分布

VBE: $$q^*(\boldsymbol{y}_i) = \prod_{j=1}^{N_i} \mathrm{Mult}(1, r_{ij1},\ldots,r_{ijK}),$$

VBM: $$q^*(\boldsymbol{\pi}_i) = \mathrm{Dir}(\alpha_{i1},\ldots,\alpha_{iK}),$$
$$q^*(\boldsymbol{\mu}_{ik},\boldsymbol{\Lambda}_{ik}) = \mathcal{N}(\boldsymbol{m}_{ik},(\beta_{ik}\boldsymbol{\Lambda}_{ik})^{-1})\mathcal{W}(W_{ik},\nu_{ik}), \forall k,$$

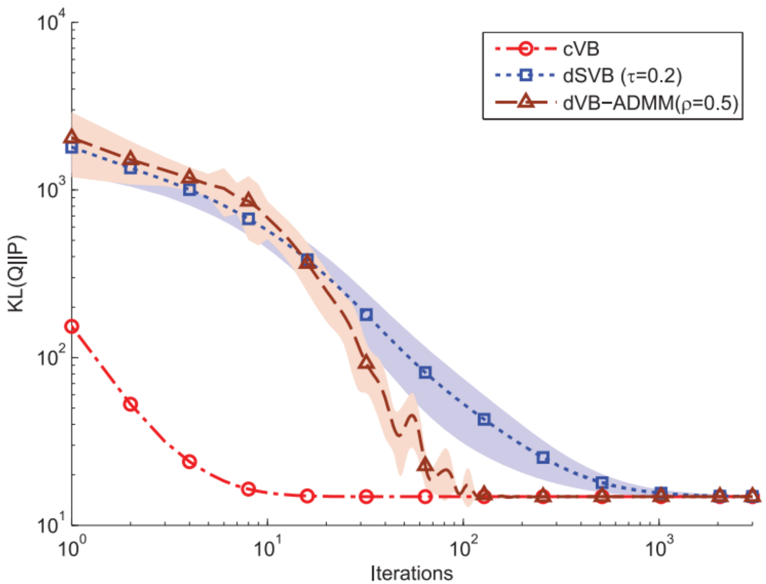
全局变分分布

- 全局自然参数向量

$$\boldsymbol{\phi}_{\boldsymbol{\pi}_i} = [\alpha_{i1}-1,\ldots,\alpha_{iK}-1]^T.$$

$$\boldsymbol{\phi}_{\boldsymbol{\mu}_{ik},\boldsymbol{\Lambda}_{ik}} = \begin{bmatrix} \frac{\nu_{ik}-D}{2} \\ -\frac{1}{2}\boldsymbol{W}_{ik}^{-1} - \frac{\beta_{ik}}{2}\boldsymbol{m}_{ik}\boldsymbol{m}_{ik}^T \\ \beta_{ik}\boldsymbol{m}_{ik} \\ -\frac{1}{2}\beta_{ik} \end{bmatrix}, \forall k = 1,\ldots,K.$$

# 仿真：验证算法有效性



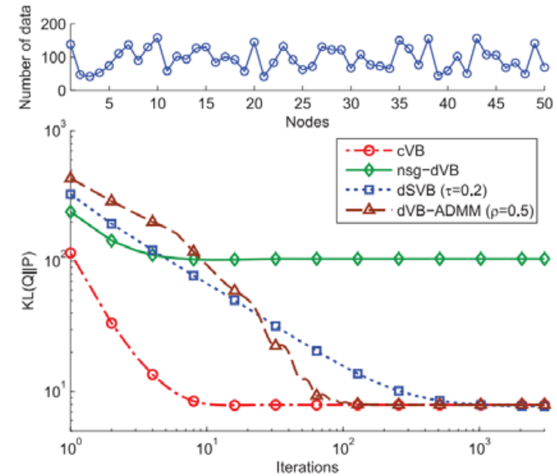




Fig. 9. The number of data at each node (top) and the average performance (bottom) of the dSVB and dVB-ADMM with the imbalanced data, compared with the nsg-dVB and cVB.

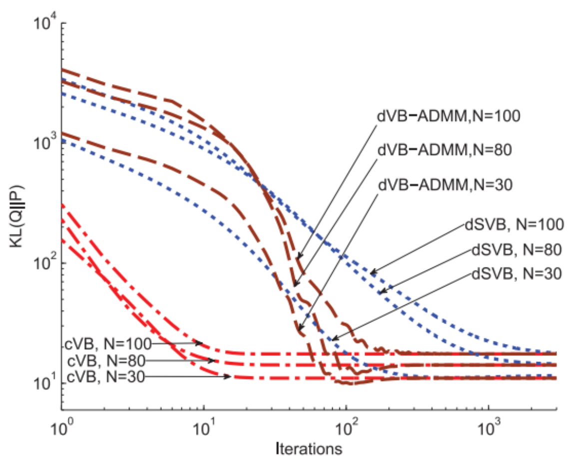



Fig. 10. The average performance of the dSVB and dSV-ADMM with different network sizes ($N = 30, 80, 100$). The tuning parameters are set as $\tau = 0.2$, $\rho = 0.5$ for all cases.

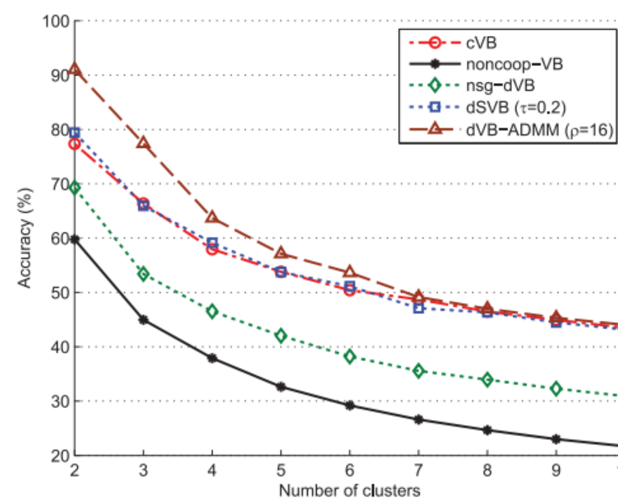



Fig. 13. Accuracy versus the number of clusters on the COIL-20 dataset.

# 其他应用



- ❑ 分布式变分贝叶斯(dVB)算法的其他应用
  - ❑ **稀疏信号重建**：分布式贝叶斯压缩感知，联合稀疏贝叶斯学习
    - 研究基于dVB的贝叶斯压缩感知算法，实现传感器网络上的联合稀疏信号重建。（进行中）
  - ❑ **目标跟踪**：卡尔曼滤波
    - 研究非高斯噪声下的状态空间模型，并用dVB算法实现在分布式网络环境下的目标跟踪。（进行中）
  - ❑ **目标识别**：贝叶斯多目标分类 ,多任务高斯过程
    - 研究基于dVB的贝叶斯多任务学习，并应用到雷达传感器网络上的目标识别。

# 研究成果或进行中工作



- **Junhao Hua**, Chunguang Li,“Distributed Variational Bayesian Algorithms Over Sensor Networks”, *IEEE Transactions on Signal Processing*, vol.64, no.3, pp.783-798, Feb. 2016.
- **Junhao Hua**, Chunguang Li, Hui-Liang Shen, “Distributed Learning of Predictive Structures from Multiple Tasks Over Networks”, *IEEE Transactions on Industrial Electronics*, preprint, Jul. 2016, DOI: 10.1109/TIE.2016.2588463.
- **Junhao Hua**, Chunguang Li, Distributed Joint Sparse Recovery with Beta Process Prior via Quantized Communication, preparing.
- **Junhao Hua**, Chunguang Li, Distributed Robust Kalman Filtering By Variational Bayesian Approximations, preparing.