Image Segmentation Using Inverted Dirichlet Mixture Model and Spatial Information

Jai Puneet Singh, Nizar Bouguila

CIISE, Concordia University, Montréal, Québec, Canada

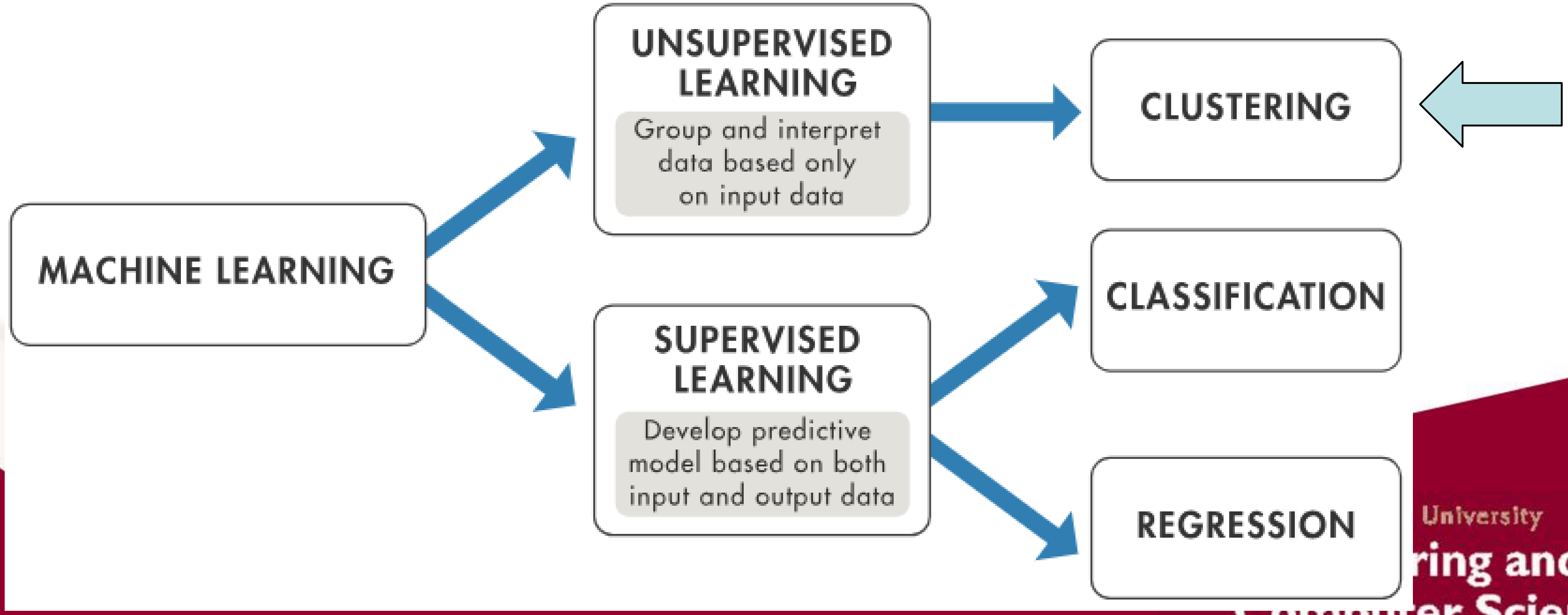
CCECE 2018

May 11, 2018

Overview

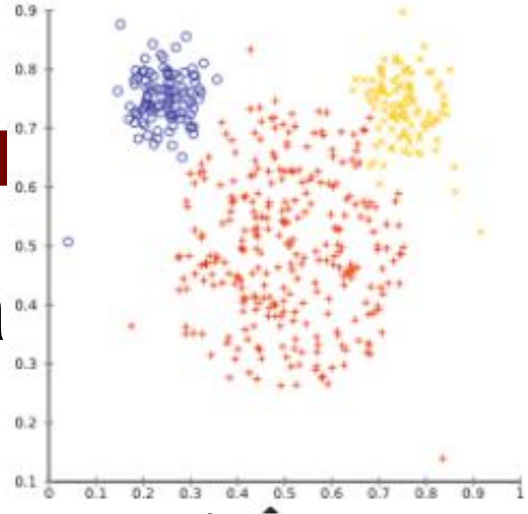
Supervised vs Unsupervised

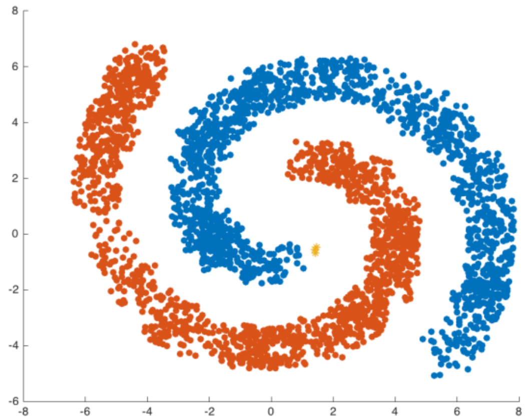
- Supervised: Could be more accurate but have learning bias, need training.
- Unsupervised: Flexible, New Patterns



What’s Machine Lea

- “Learn” information from data
  - Pattern Recognition
    - Distance-based: K-means,…
    - Connectivity-based: DB-scan,…
    - Probability-based
    - …….

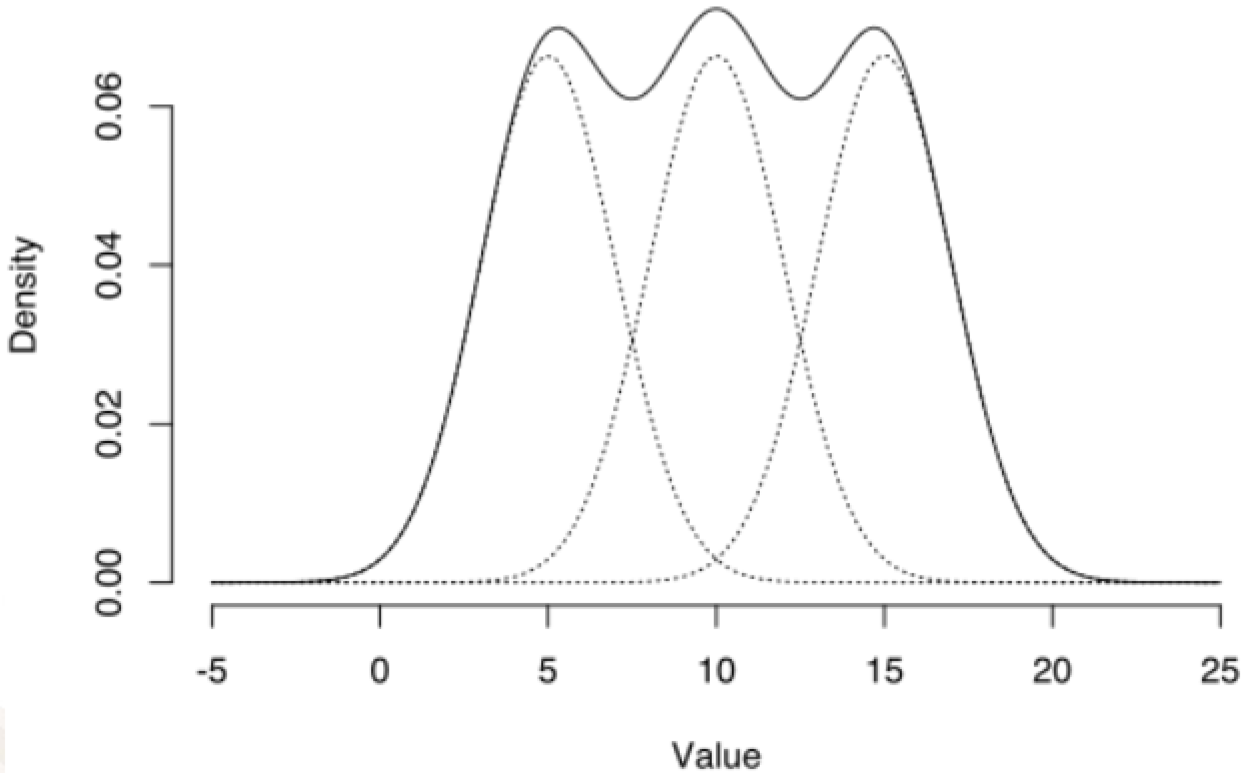




Concordia University
Engineering and
Computer Science

Mixture Model

- Respect dependences between clusters



Related Works

K-Means Clustering Approaches

It is partition based approach and uses distance metrics to find the nearest neighbors. The objective function of K-means can be represented as follows:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^j - c_j \right\|^2 \tag{1}$$

1. **L1 Distance**: Kashima et al. used distance on proportional data.
2. **Kullback-Leibler Divergence**: It's used for different distributions

Proposed Method

Using distance metrics in K-means

Algorithm 1 K-Means Algorithm

1: Set the Initial number of centroids randomly or sequentially
2: Calculate the distance between each data point and cluster centers
3: **repeat**:
4: Assign the minimum **distance data points** to cluster center whose distance is minimum to that point.
5: Recalculate the cluster center using:
6: $c_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x(i)$; m_i represents total number of data points in (i) cluster
7: Re-calculate the distance between each data point and newly obtained cluster center
8: **until** : No data point is reassigned.

Proposed Method

Using distance metrics in K-means

S.No.	Distance Name	Distance Metrics
1	Euclidean Distance	$d_E^2(x,y) = \sum_i (x_i - y_i)^2$
2	Euclidean Distance log transformed Data	$d_{EL}^2(x,y) = \sum_i (\log x_i - \log y_i)^2$
3	J-divergence	$d_{jd}^2(x,y) = \sum_i (\log x_i - \log y_i)(x_i - y_i)$
4	Jeffery's-Matusita Distance	$d_m^2(x,y) = \sum_i \left(\sqrt{x_i} - \sqrt{y_i}\right)^2$
5	Manhattan Distance (L1 Distance)	$d_{L1}^2(x,y) = \sum_i \lvert x_i - y_i \rvert$
6	**Aitchison's Distance**	$d_{AD}(x,y) = \frac{1}{D} \sum_{i<j} \left(\log \frac{x_i}{x_j} - \log \frac{y_i}{y_j}\right)$ $d_{AD}^2(x,y) = \sum_{k=1}^{D} \left(\log \frac{x_i}{g\left(x_j\right)} - \log \frac{y_i}{g\left(y_j\right)}\right)$
7	Cosine Distance	$d_C(x,y) = 1 - \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}$
8	Mahalonbis Distance	$d_m^2(x,y) = (x-y)^T S^{-1} (x-y)$

Markov Random Field

1. The Markov Random Field distribution can be shown as follow:

$$f(\Pi) = Z^{-1} \exp\left\{-\frac{1}{T} U(\Pi)\right\} \tag{2}$$

where Z is the normalizing constant and T is the temperature constant. $U(\Pi)$ is the smoothing prior.

Choice of distribution

1. The **Guassian distribution**, though widely used and popular, it has the following drawbacks;
   - Symmetric structure.
   - Not suitable with skewed data or data generated from non-Guassian sources.
2. The **Inverted Dirichlet distribution**, though it is constrained on a simplex, it is very flexible and can assume several shapes depending on its parameter.
   - The drawback is the proper initialization of the data.
   - Popularly, it is done by K-means method.
3. Other models explored in mixture modeling are inverted Dirichlet, generalized Dirichlet, scaled Dirichlet etc.

Proposed Method

Mixture Models

1. We have used Inverted Dirichlet Mixture Model which can be shown as:

$$p\left(\vec{X}_i|\vec{\alpha}_j\right) = \frac{\Gamma|\vec{\alpha}_j|}{\prod_{d=1}^{D+1}\Gamma\left(\alpha_{jd}\right)}\prod_{d=1}^{D}X_{id}^{\alpha_{jd}-1}\left(1+\sum_{d=1}^{D}X_{id}\right)^{-|\vec{\alpha}|} \tag{3}$$

where $X_{id} > 0$, $d = 1, 2, ..., D$, $X_{i1} + X_{i2}, ... + X_{iD} = 1$, $\vec{\alpha}_{jd} = (\alpha_{j1}, \alpha_{j2}, ..., \alpha_{jD})$, $|\vec{\alpha}_j| = \sum_{d=1}^{D+1}\alpha_{jd}$ and $\alpha_{jd} > 0$ represents parameter vector for j^{th} component.

Proposed Method

Integration of MRF with Inverted Dirichlet distribution

The MRF can have different types of models depending upon the energy $U(\Pi)$.

1. In Bayesian auto logistic model, the $U(\Pi)$ is chosen to incorporate the spatial information as:

$$U(\Pi) = \sum_{i=1}^{N}\sum_{j=1}^{K}\alpha_{ij}\pi_{ij} + \sum_{i=1}^{N}\sum_{j=1}^{K}\sum_{m\in\delta_i}\beta_{ijm}\pi_{ij}\pi_{mj} \tag{4}$$

2. Other method used by keeping β as a constant

$$U(\Pi) = \beta\sum_{i=1}^{N}\sum_{m\in\delta_i}\left[1+\left(\sum_{j=1}^{K}\left(\sum_{j=1}^{K}(\pi_{ij}-\pi_{mj})\right)\right)\right] \tag{5}$$

Proposed Method

Integration of MRF with Dirichlet distribution

3. Keeping the constant value for β and setting Z=1 and T=1.

$$U(\Pi) = \beta \sum_{i=1}^{N} \sum_{m \in \delta_i} \sum_{j=1}^{K} (\pi_{ij} - \pi_{mj})^2 \tag{6}$$

4. The following smoothing prior is used:

$$U(\Pi) = -\sum_{1}^{N} \sum_{j=1}^{M} G_{ij}^{(t)} \log p_{ij}^{(t+1)} \tag{7}$$

Proposed Method

Integration of MRF with Dirichlet distribution

The joint conditional density of data can be presented as:

$$f\left(\mathcal{X}|\theta\right) = \prod_{i=1}^{N} p\left(\vec{X}_i|\theta\right) = \prod_{i=1}^{N}\sum_{j=1}^{M} p_j p\left(\vec{X}_i|\vec{\alpha}_j\right) \tag{8}$$

The Bayes rule for posterior probability can be represented as:

$$f\left(\theta|\mathcal{X}\right) \propto f\left(\mathcal{X}|\theta\right) f\left(\Pi\right) \tag{9}$$

Proposed Method

Integration of MRF with Dirichlet distribution

By Considering the Bayes rule, we will find log-likelihood function which can be derived as:

$$L\left(f\left(\theta|\mathcal{X}\right)\right) = \sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{K} p_j p\left(\vec{X}|\vec{\alpha}\right) \right\} - \log Z - \frac{1}{T} U\left(\Pi\right) \quad (10)$$

Maximizing equation (11) we get the expanded equation with the hidden variable z_{ij}

$$L\left(f\left(\theta|\mathcal{X}\right)\right) = \sum_{i=1}^{N} \sum_{j=1}^{M} z_{ij}^{t} \left\{ \log p_{ij}^{(t+1)} + \log p\left(\vec{X}_i|\vec{\alpha}_j\right) \right\} - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{M} G_{ij}^{t} \log p_{ij}^{(t+1)} \quad (11)$$

Proposed Method

Integration of MRF with Dirichlet distribution

Using the novel method introduced by Nguyen et al which is:

$$G_{ij}^{t} = \exp\left[\frac{\beta}{2N_i}\sum_{m\in\delta_i}\left(z_{mj}^{(t)} + p_{mj}^{(t)}\right)\right] \tag{12}$$

Using the discussed methods the prior probability at $(t+1)$ becomes :

$$p_{ij} = \frac{z_{ij}^{(t)} + G_{ij}^{(t)}}{\sum_{m=1}^{k}\left(z_{im}^{(t)} + G_{ik}^{(t)}\right)} \tag{13}$$

Proposed Method

EM Algorithm Dirichlet Mixture Model with MRF

Algorithm 1 EM Algorithm Inverted Dirichlet Mixture Model with MRF

Apply K-means on image data points to obtain initial k clusters for segmentation.
The initial estimates for each mixture component j:
$\alpha_{jD+1} = \frac{E(X_d)^2 + E(X_d)}{Var(X_d)} + 2$ and $\alpha_{jd} = E(X_d)(\alpha_{jD+1} - 1), d = 1, \ldots, D.$
Initialization of above equation using Method of Moments as proposed by the author in [12] to obtain the two parameters.
Use the image data points to update the mixture parameters.
E-Step: Compute the posterior probability $z_{ij}^{(t)}$
M-Step:
repeat:
    Update priors p_j using equation 15 .
    Update the parameters $\vec{\alpha}$ using Newton Raphson method [16].
until : $p_j \leq \epsilon$, discard j and go to E-Step.

Experimental Results

Normalized Probabilistic Rand

$$\text{NPR Index} = \frac{\text{PR Index - Expected Index}}{\text{Maximum Index - Expected Index}} \tag{14}$$

The Expected value of PR Index can be given as follow:

$$E\left[PRI\left(S_{test}, \{S_k\}\right)\right] = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i<j}} \left[p'_{ij} p_{ij} \left(1 - p'_{ij}\right) \left(1 - p_{ij}\right)\right] \tag{15}$$

Experimental Results

Normalized Probabilistic Rand

1. We have used Berkeley Segmentation Data-set 500 (BSDS500) which is an extension of Berkeley Segmentation Data-set 300.

	DMM	IDMM	FRDMM	FRIDMM
NPR Index Sample Mean	0.4247	0.4390	0.6182	0.6793

Table: NPR index sample for Gaussian Mixture model (GMM), Dirichlet Mixture model (DMM), Fast and Robust Gaussian mixture model (FRGMM) and Fast and Robust Dirichlet mixture model (FRDMM)

Experimental Results

Segmented Images

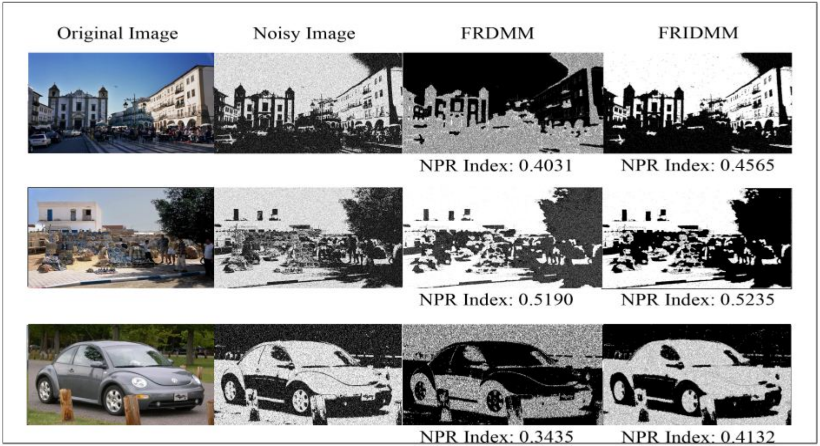


Figure: Segmentation of images from Berkeley 500 database

Conclusion

- We have proposed integration of MRF into Inverted Dirichlet Mixture model.
- We have earlier did the analysis which showed us that initialization by K-means using Aitchison's distance on proportional data improves the accuracy of the model.
- The NPR Sample mean has been calculated of image data sets, hence showing the improvement with our approach.

Thank You