Tutorial on spectral clustering (1)   Wednesday, June 15, 2011

Refs: (1) U. von Luxburg, A tutorial on spectral clustering, Stat. Comp, 2007.

(2) Shi and Malik, Normalized cuts and image segmentation, PAMI, 2000

Starting point: Undirected graph $G=(V,E)$, $V=\{v_1,\dots,v_n\}$.
with weighted edges $w_{ij}=w_{ji}\geq 0$. [similarities, large w_{ij} means strong bond].

Degree of a vertex v_i : $d_i=\sum_{j=1}^{n} w_{ij}$, $D=\text{diag}(d_i)$ is the degree matri

For two not necessarily disjoint subsets of V, $W(A,B)=\sum_{i\in A, j\in B} w_{ij}$.

Size of a subset $A\subset V$:

$|A|$ = number of vertices in A (un-weighted)

$\text{vol}(A)=\sum_{i\in A} d_i$ (weighted volume).

From similarities to graphs: $(x_1,\dots,x_n)$ data points with similarities $s_{ij}\geq 0$.

- ε-neighborhood graph: $w_{ij}=\begin{cases}1, & \text{if } s_{ij}>T_\varepsilon.\\ 0, & \text{otherwise.}\end{cases}$
- K-nn graph: $w_{ij}=\begin{cases}s_{ij}, & \text{if } v_i \text{ is a } K\text{-nn neighbor of } v_j \text{ or vice versa}\\ 0, & \text{otherwise}\end{cases}$
- fully connected graph: $w_{ij}=s_{ij}$

Example (Shi & Malik): V: set of pixels in an image

$$w_{ij}=e^{-\frac{\|I_i-I_j\|^2}{\sigma_I^2}}\cdot\begin{cases}e^{-\frac{\|x_i-x_j\|^2}{\sigma_x^2}}, & \text{if } \|x_i-x_j\|<r\\ 0, & \text{otherwise.}\end{cases}$$

Tutorial on spectral clustering (2)

June 15, 2011

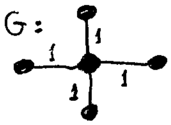
The Graph Laplacian matrix

(ref. Chung, Spectral graph theory, AMS-1997)

Unnormalized graph Laplacian:

$$L = D - W \quad \left[\begin{array}{l}\text{does not depend}\\ \text{on } w_{ii} \text{ - cancel out}\end{array}\right]$$

[Why "Laplacian"? :



, $L = \begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}$

which is the standard discretization for the (negative) Laplacian operator $-\nabla^2 u = -(u_{xx} + u_{yy})$.]

Properties of L ($f \in \mathbb{R}^n$):

(1) $f^T L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} (f_i - f_j)^2$

(2) L is symmetric pos. semi-def.

(3) smallest eigenvalue is 0, eigenvector $\mathbb{1}$.

(4): $0 = \lambda_1 \le \lambda_2 \le \dots \le \lambda_n$.

(5): If $0 = \lambda_1 = \lambda_2 = \dots = \lambda_k < \lambda_{k+1} \le \dots \le \lambda_n$, then G has k-connected components and $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_k}$ are eigenvectors.

Normalized graph Laplacians:

$$L_{sym} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2} \quad \text{(Ng, Jordan, Weiss, 2002)}$$

$$L_{rw} = D^{-1} L = I - D^{-1} W \quad \text{(Shi \& Malik, 2000)}$$

Properties:

(1) $f^T L_{sym} f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} \left(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right)^2$

(2) (λ, u) eigenpair of L_{rw} $\Leftrightarrow$ (λ, w), $w = D^{1/2} u$, eigenpair of L_{sym}.

(3) (λ, u) eigenpair of L_{rw} : $Lu = \lambda D u$ (generalized eigenproblem)

(4) $(0, \mathbb{1})$ eigenpair of L_{rw}, $(0, D^{1/2}\mathbb{1})$ eigenpair of L_{sym}.

(5) L_{sym}, L_{rw} are positive semi-def. and $0 = \lambda_1 \le \dots \le \lambda_n$.

(6) If $0 = \lambda_1 = \dots = \lambda_k < \lambda_{k+1} \le \dots \le \lambda_n$, then G has k-connected components

The eigenspace of $\lambda = 0$ is spanned by $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_k}$ for L_{rw} and $D^{1/2}\mathbb{1}_{A_i}$ for L_{sym}

Tutorial on spectral clustering. (3)

June 15, 2011

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number of clusters k.

(1) Construct $G=(V,E)$ and W from S.

(2) Compute the un-normalized Laplacian $L=D-W$

(3) Compute the first k eigenvectors of L

or Solve the general. eigenvalue prop. $Lu=\lambda Du$ (for L_{rw})

or Compute the first k eigenvectors of $L_{sym}=D^{-1/2} L D^{-1/2}$ (for L_{sym}).

(4) Let $U=[u_1,\dots,u_k] \in \mathbb{R}^{n \times k}$

$$= \begin{bmatrix} y_1^T \\ \vdots \\ y_n^T \end{bmatrix}, \quad y_i \in \mathbb{R}^k$$

(5) Cluster the points $\{y_i\}$, $i=1:n$ using, e.g. k-means, into clusters $C_{1:k}$.

Output Clusters $A_1,\dots,A_k$, with $A_i=\{j \mid y_j \in C_i\}$.

Main point: In the new representation $\{y_i\}$ clustering is much easier.

Graph cut point of view.

$$\mathrm{cut}(A_1,\dots,A_k) = \frac{1}{2}\sum_{i=1}^{k} W(A_i,\bar{A}_i)$$

$$\text{Ratio-cut}(A_1,\dots,A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{W(A_i,\bar{A}_i)}{|A_i|} = \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i,\bar{A}_i)}{|A_i|}$$

$$\mathrm{Ncut}(A_1,\dots,A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{W(A_i,\bar{A}_i)}{\mathrm{vol}(A_i)} = \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i,\bar{A}_i)}{\mathrm{vol}(A_i)}$$

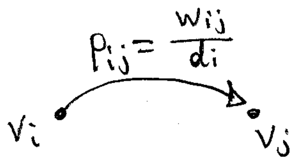
Relaxations:

Ratio-cut ⇝ unnormalized spectral clustering

Ncut ⇝ normalized spectral clustering of Shi & Malik

Tutorial on spectral clustering (4)

Random walks:



Unique stationary distribution if graph connect and non-bipartite.
$\pi = (\pi_1, \dots, \pi_n)^T$, with $\pi_i = \frac{d_i}{vol(V)} = \frac{d_i}{\sum_{i=1}^{n} d_i}$

Transition matrix: $P = D^{-1} W$

(note that $L_{RW} = I - P$). ⤳ (λ, u) eigenpair of L_{rw} ⇔ $(1-\lambda, u)$ eigenpair of P
smallest eigenvectors of L_{rw} ↔ largest eigenvectors of P.

$N_{cut}(A, \bar{A}) = P(\bar{A} | A) + P(A | \bar{A})$, if we start the random walk from $X_0 = \pi$.

where $P(B | A) = P(X_1 \in B \mid X_0 \in A)$

Commute distance (resistance distance). [ref. Qiu & Hancock, PAMI-07].

C_{ij} : [expected time it takes the random walk to travel from vertex v_i to vertex v_j and back]

⤳ [instead of looking for the single shortest path it takes into account several reasonably short paths.].

⤳ also see L. Grady, Random walks for image segmentation, PAMI-06.

$$C_{ij} = vol(V) \left(l^{+}_{ii} - 2 l^{+}_{ij} + l^{+}_{jj} \right) = vol(V) (e_i - e_j)^T L^{+} (e_i - e_j),$$

where $L^{+} = pinv(L)$

⤳ $\sqrt{C_{ij}}$ can be considered as Euclidean distance on the vertices of the graph.

⤳ Construct an embedding which maps the vertices v_i of the graph on points $z_i \in \mathbb{R}^n$ such that the Euclidean distances between the points z_i coincide with the commute distances on the graph.