Lecture 10: PGM — Structure Estimation

Qinfeng (Javen) Shi

13 Oct 2014

Intro. to Stats. Machine Learning
COMP SCI 4401/7401

Table of Contents I

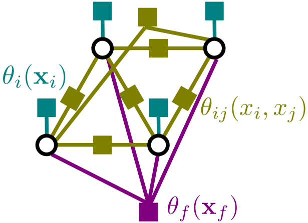
How to get the graph at the first place?

- Human heuristics
- Learn graphs from the data
  - Learn from labels (statistical independence testing or mutual information)
  - Learn from both labels and features (omitted)
- Infer the graph and labels jointly.

Manually specify a fixed graph

(a) original image

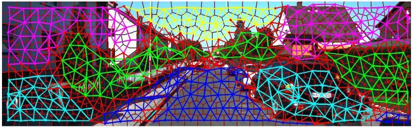
(b) segmented image



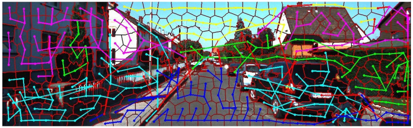
(c) graph structure

Use a simple rule

(d) original image



(e) graph via super-pixel **adjacency**



(f) graph via **distance mst**

Learn graphs from data

Assumptions:

1. the unknown underlying graph is fixed;
2. training data are samples from the distribution represented by the underlying graph;
3. the number of nodes (*i.e.* variables) is known in advance, and the edges are unknown.
4. can extend to multiple fixed underlying graphs, however, each graph shall have enough samples.

Learn graphs from labels only

Techniques:

- statistical independence testing
- mutual information (*e.g.* ChowLiu Tree algorithm)
- ...

Problem:

- it only considers labels (output), and does not consider features (input)
- it does not consider label cost functions

Learn graphs from both labels and features

Idea:

- One can enforce sparsity (*e.g.* by $\| \mathbf{w} \|_1$ regulariser) in structured SVM or CRFs (Lecture 9) to achieve a sparse $\mathbf{w}$.
- If certain block of $\mathbf{w}$ being zero or non-zero corresponds to existence of edge, learning such $\mathbf{w}$ is learning edges.

Infer graphs and labels jointly

Assumptions:

1. the unknown underlying graph is ~~fixed~~ [changing];
2. training data are samples from the distribution represented by the underlying graph;
3. the number of nodes (*i.e.* variables) is known in advance, and the edges are unknown.
4. can extend to multiple fixed underlying graphs, however, each graph ~~shall~~ [does not] have enough samples.

$\Rightarrow$Not enough samples to learn the graphs.

Infer graphs and labels jointly

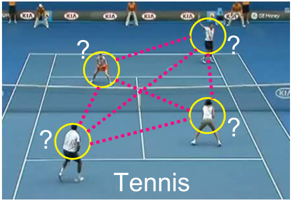


Figure : Unknown MRF graphs and unknown labels

$G = (V, E)$, where V node set is known, and E edge set is unknown. To find the best label and the best E jointly,

$$(\mathbf{y}^*, E^*) = \operatorname*{argmax}_{\mathbf{y} \in \mathcal{Y}, E \in \mathcal{E}} \sum_{i,j \in E} \theta_{ij}(y^{(i)}, y^{(j)}) + \sum_{i \in V} \theta_i(y^{(i)}). \qquad (1)$$

Alternating method

Lan etal (NIPS 2010) alternates between solving $\mathbf{y}$ and E.

Initialise $\mathbf{y}_1$ randomly.
for $t = 1$ **to** T **do**

$$E_t = \operatorname*{argmax}_{E \in \mathcal{E}} \sum_{i,j \in E} \theta_{ij}(y_t^{(i)}, y_t^{(j)}), \tag{2}$$

$$\mathbf{y}_{t+1} = \operatorname*{argmax}_{\mathbf{y} \in \mathcal{Y}} \sum_{i,j \in E_t} \theta_{ij}(y^{(i)}, y^{(j)}) + \sum_{i \in V} \theta_i(y^{(i)}). \tag{3}$$

end for
$G = (V, E_T)$.

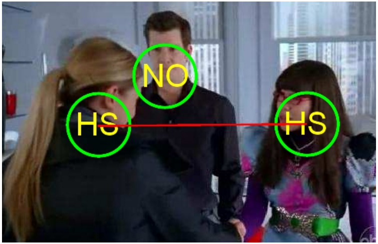
Bilinear formulation

Wang etal (CVPR2013) introduces bilinear program (BLP) form.

$$\max_{\{z_{ij}\},\{\mu_{i,j},\mu_i\}} \sum_{i,j\in V} \sum_{y^{(i)},y^{(j)}} \theta_{i,j}(y^{(i)},y^{(j)})\mu_{i,j}(y^{(i)},y^{(j)})z_{ij}$$
$$+ \sum_{i\in V} \sum_{y^{(i)}} \theta_i(y^{(i)})\mu_i(y^{(i)}) \tag{4}$$
$$\text{s.t. } \sum_{y_i} \mu_{i,j}(y^{(i)},y^{(j)}) = \mu_j(y^{(j)}), \quad \sum_{y^{(j)}} \mu_{i,j}(y^{(i)},y^{(j)}) = \mu_i(y^{(i)}),$$
$$\sum_{y^{(i)}} \mu_i(y^{(i)}) = 1, \quad \mu_{i,j}(y^{(i)},y^{(j)}) \geq 0, z_{ij} = z_{ji}, z_{ij} \in [0,1],$$
$$\forall i,j \in V, y^{(i)}, y^{(j)}.$$

Inference with unknown graphs

Alg.	MCSVM					SSVM					Lan's					BLP				
A/A	**NO**	**HS**	**HF**	**HG**	**KS**	**NO**	**HS**	**HF**	**HG**	**KS**	**NO**	**HS**	**HF**	**HG**	**KS**	**NO**	**HS**	**HF**	**HG**	**KS**
NO	0.37	0.07	0.21	0.11	0.24	0.20	0.40	0.27	0.06	0.06	0.11	0.36	0.19	0.20	0.13	**0.49**	0.20	0.13	0.13	0.05
HS	0.01	0.55	0.06	0.17	0.21	0.10	0.51	0.21	0.11	0.06	0.09	0.52	0.14	0.15	0.10	0.18	**0.56**	0.09	0.08	0.09
HF	0.09	0.03	0.52	0.21	0.14	0.08	0.11	0.61	0.08	0.12	0.02	0.14	0.58	0.18	0.08	0.11	0.09	**0.63**	0.07	0.10
HG	0.02	0.14	0.20	0.49	0.15	0.05	0.15	0.11	0.58	0.11	0.03	0.06	0.11	0.55	0.26	0.03	0.10	0.10	**0.70**	0.06
KS	0.07	0.11	0.09	0.05	0.67	0.02	0.26	0.14	0.12	0.46	0.01	0.07	0.15	0.11	0.67	0.06	0.08	0.04	0.13	**0.69**



That's all

Thanks!