Computer Vision II

Graphical models: statistical inference and learning

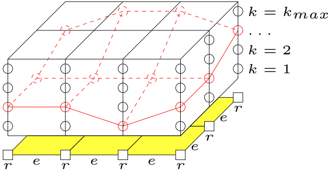
Carsten Rother, Dmitrij Schlesinger

SS2016, 15.07.2016

COMPUTER VISION LAB DRESDEN

TECHNISCHE UNIVERSITÄT DRESDEN

Graphical models (reminder)



$G(V, \mathcal{E})$ – basis graph,
$i \in V$ – nodes, $(i, j) \in \mathbb{E}$ – edges
$x : V \to C$ – an image,
$x_i \in C$ – image color at position i
L – label set (e.g. the set of segments)
$y : V \to L$ – labeling,
$y_i \in L$ – a label at position i

There is **energy** $E(x, y)$ associated with each pair (x, y)
Special case: pairwise models

$$E(x, y) = \sum_{i \in V} \phi_i(x, y_i) + \sum_{(i,j) \in \mathcal{E}} \phi_{ij}(x, y_i, y_j),$$

with $\phi_i(x) : L \to \mathbb{R}$ and $\phi_{ij}(x) : L \times L \to \mathbb{R}$

Graphical models (reminder)

The posterior probability distribution (consider CRF) is

$$p(y|x) = \frac{1}{Z(x)} \exp\Big[-E(x,y)\Big]$$

Inference: observe x, predict y

An "obvious" choice – take the most a-posteriori probable y

$$y^* = \arg\max_y p(y|x) = \arg\min_y E(x,y)$$

(MAP), i.e. an Energy Minimization problem

Are there alternatives ?
Is MAP reasonable for structured prediction at all ?

A motivating example

	Q_1	Q_2	$\dots$	Q_n
P_1	1	0	$\dots$	1
P_2	0	1	$\dots$	0
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
P_m	0	1	$\dots$	0
"$\sum$"	?	?	$\dots$	?

Consider a "questionnaire": m persons answer n questions. Furthermore, let us assume that persons are rated – a "reliability" measure is assigned to each one.

The goal is to find the "right" answers for all questions.

Strategy 1:
Choose the **best** person and take **all** his/her answers.

Strategy 2:

- Consider a particular question
- Look, what **all** the people say concerning this, do (weighted) voting

A motivating example, interpretation

"People" are labelings y,
the "reliability measure" is the posterior $p(y|x)$

"Questions" are pixels i,
one "elementary answer" is a possible label at this pixel

The "Strategy 1" is MAP

The "Strategy 2" can be derived within **Bayesian Decision Theory** using so-called **additive cost-functions**
(see "Machine Learning" lectures).

Maximum Marginal Strategy

1. Compute **marginal** probability distributions of labels

$$p(y_i{=}l|x) = \sum_{y:y_i=l} p(y|x)$$

   for each variable i and each label value l

2. Decide for each variable "independently" according to its marginal probability distribution

$$y_i^* = \arg\max_l p(y_i{=}l|x)$$

There are also other marginal-based strategies depending on the nature of the label set

Minimum Marginal Squared Error (MMSE)

Example – stereo, the labels represent depth-values

The strategy is the average

$$y_i^* = \sum_l l \cdot p(y_i{=}l|x)$$

MAP

MMSE

Statistical learning (very short)

- Maximum Likelihood:
  maximize the probability of the given training sample
- **Key observation**: the energy can be written as a dot-product (see "Optimization for Machine Learning"):

$$E(x, y) = \langle \psi(x, y), w \rangle$$

  hence, MRF-s are members of the **Exponential Family**
- See "Machine Learning" again how to learn EF-models

At the end some marginal probability distributions are necessary for both inference and learning (gradient of the likelihood). It is a very hard problem :-(. Approximate solutions: sampling techniques, mean-field approximations, some upper bounds ...

Take home messages

Statistical treatment of structured models is extremely powerful !!!

Inference: large spectrum of different decision strategies for the same statistical model

The design of an appropriate decision strategy is as important as the design of the appropriate model

Learning: well founded principles

Both statistical inference and learning are computationally extremely hard