Higher-order Statistics

- Standard CRF models usually trained using simple low-order losses.
- In real-world often more complex higher-order losses used for evaluation.
- Goal here: Train classifier directly with this higher-order loss.
- Our work introduces a higher-order loss for which we can train structured SVMs exactly.

Model

Train a predictor of the form

$$f_w(x) = \arg\min_{y \in Y} E(y, x, w).$$

$$E(y, x, w) = -\langle w, \phi(x, y) \rangle = \sum_{i \in V} \psi_i(y_i, x, w^i) + \sum_{(i,j) \in E} \psi_{ij}(y_i, y_j, x, w^{ij}).$$

Max-margin Learning

The structured SVM considers the following quadratic program:

$$\min_{w} \frac{\lambda}{2} ||w||^2 + \sum_{y \in Y} c^n |

s.t. \max_{y \in Y} [\langle w, \phi(x^n, y) \rangle + \Delta_y(y)] - \langle w, \phi(x^n, y^*) \rangle \geq c^n \quad \forall n \geq 0.$$

- Optimizes convex upper bound on classification error.
- Loss $\Delta_y(y)$ measures how bad it is to predict y instead of $y^*$.
- Solved by the cutting planes algorithm.
- Line 5: Loss augmented inference.

Loss Augmented Inference

- Need to efficiently solve the problem:
  $$\min_y E(y, x, w) - \Delta_y(y).$$
- Notice the negative sign!
- We assume $y_i$ is binary and $E(y, x, w)$ is submodular. Therefore: energy minimization in the original model is exactly solvable.

Loss Functions

- Should reflect scoring used for evaluation.
- But at the same time loss augmented inference should also be tractable!
- In practice for many segmentation problems Hamming loss is used:
  $$\Delta_{\text{Hamming}}(y) = \sum_{i \in V} y_i \neq y_i^*.$$

Loss augmented inference has same complexity as inference for original model.
- Only modifies the unaries.
- A low-order loss. What about higher-order losses?
- Here we study the label-count loss:
  $$\Delta_{\text{count}}(y) = \sum_{i \in V} y_i - \sum_{i \in V} y_i^*.$$

Useful if we are only interested in predicting the number of foreground pixels, but not their location.
- Unfortunately label-count loss no longer factorizes!

Lower and Upper Envelopes

- Many higher order functions can be represented as:
  $$f^\Delta(y) = \Delta_{\text{max}} \circ f^\text{min}(y)$$

where $\Delta = \{ \text{max}, \text{min} \}$, and $\text{Q}$ indexes a set of linear functions.
- min: lower envelope, max: upper envelope.
- Inference for upper envelope substantially more difficult (min-max).
- Label-count is upper envelope representable.
- Fortunately, negative sign makes loss lower envelope representable.

$$\Delta_{\text{count}}^\Delta(y) = \sum_{i \in V} y_i - \sum_{i \in V} y_i^*$$

Label-count Loss Augmented Inference

Obtain the pairwise minimization problem:

$$\min_{y \in \{0,1\}} E(y, x, w) + 2 \sum_{i \in V} y_i - \sum_{i \in V} y_i^* + \sum_{i \in V} y_i - \sum_{i \in V} y_i^*.$$

- Can be solved by standard graph-cut with an auxiliary variable.
- Or alternatively by two standard graph-cut calls with modified unaries.
- Former approach also works if label-count loss for several parts is desired.
- Iterative breadth-first search graph-cut leads to better performance.

Cell Segmentation

- Goal: Counting number of mitochondria cell pixels in an electrosopic image.
- Right: Hamming loss trained model minus count-loss trained model.

Background-Foreground Segmentation

Conclusions

- Max-margin learning with the label-count loss can be done exactly.
- Leads to better results if only interested in the number of foreground pixels.
- Also see Danny Tarlow’s poster here at AISTATS.

References