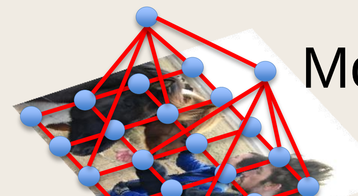
Empirical Minimum Bayes Risk Prediction: How to extract an extra few % performance from vision models with just three more parameters

Vittal Premachandran (NUS), Daniel Tarlow (Microsoft Research), Dhruv Batra (Virginia Tech)

VirginiaTech

Overview of EMBR Predictor

$\mathbf{x}^{\text{test}}$



Model

$P(\mathbf{y}|\mathbf{x}^{test}) = \frac{1}{Z} e^{S(\mathbf{y};\mathbf{x}^{test})}$

Loss Function $\ell(\mathbf{y}, \hat{\mathbf{y}})$ e.g. $\frac{\cap}{\cup}$

Classical MBR

Prediction / Decision

$$\underset{\hat{\mathbf{y}}}{\text{argmin}} \sum_{\mathbf{y}} \mathbf{P}(\mathbf{y}|\mathbf{x}^{\text{test}}, \mathbf{w}^*)\ell(\mathbf{y}, \hat{\mathbf{y}})$$

Black Box

$$\underset{i \in [M]}{\text{argmin}} \sum_{j \in [M]} \tilde{\mathbf{P}}(\mathbf{y}^j|\mathbf{x}^{\text{test}}, \mathbf{w}^*)\tilde{\ell}(\mathbf{y}^j, \mathbf{y}^i)$$

λ M T

Empirical MBR

$\mathbf{y}^1$ $\mathbf{y}^2$ $\mathbf{y}^3$ $\mathbf{y}^M$

Algorithm

Algorithm 1 Empirical Minimum Bayes Risk Prediction

Input: Score function $S(\mathbf{y}; \mathbf{x})$, loss $\tilde{\ell}(\cdot, \cdot)$.
Input: Validation-selected parameters M, T, λ.
{DivMBest}
for $m \in 1, \dots, M$ **do**
  $S^m_\Delta(\mathbf{y}) \leftarrow S(\mathbf{y}) + \sum_{u \in \mathcal{V}} \sum_{m'=1}^{m-1} \lambda \cdot [\![y_u \neq y_u^{m'}]\!]$
  $\mathbf{y}^m \leftarrow \text{argmax}_{\mathbf{y}} S^m_\Delta(\mathbf{y}; \mathbf{x})$
end for
{Scores to Probabilities}
for $i \in 1, \dots, M$ **do**
  $\tilde{\mathrm{P}}(\mathbf{y}^i|\mathbf{x}) \leftarrow \frac{\exp\{\frac{1}{T} S(\mathbf{y}^i; \mathbf{x})\}}{\sum_{j=1}^{M} \exp\{\frac{1}{T} S(\mathbf{y}^j; \mathbf{x})\}}$
end for
{Prediction}
$i^* = \text{argmin}_{i \in [M]} \sum_{j \in [M]} \tilde{\mathrm{P}}(\mathbf{y}^j|\mathbf{x})\tilde{\ell}(\mathbf{y}^j, \mathbf{y}^i)$
return $\mathbf{y}^{i^*}$

Three Parameters

λ Diversity Parameter

M # Diverse Solutions

T Temperature Parameter

Qualitative Results

MAP Predictor

Score Function: $S(\mathbf{y}; \mathbf{x})$

MAP Predictor:

$$\mathbf{y}^{MAP} = \arg\max_{\mathbf{y} \in \mathcal{Y}} S(\mathbf{y}; \mathbf{x})$$

MBR Predictor

Bayes Risk: $\text{BR}(\hat{\mathbf{y}}) = \mathbb{E}_{P(\mathbf{y}|\mathbf{x})}[\ell(\mathbf{y}, \hat{\mathbf{y}})] = \sum_{\mathbf{y} \in \mathcal{Y}} P(\mathbf{y}|\mathbf{x})\ell(\mathbf{y}, \hat{\mathbf{y}})$

Gibbs Distribution: $P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} e^{S(\mathbf{y};\mathbf{x})}$

MBR Predictor:

$$\mathbf{y}^{MBR} = \arg\min_{\hat{\mathbf{y}} \in \mathcal{Y}} \text{BR}(\hat{\mathbf{y}})$$

EMBR Predictor

$$\mathbf{Y} = \{\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^M\}$$

$$\tilde{P}(\mathbf{y}|\mathbf{x}) = \frac{\exp \frac{1}{T} S(\mathbf{y}; \mathbf{x})}{\sum_{\mathbf{y}' \in \mathbf{Y}} \exp \frac{1}{T} S(\mathbf{y}'; \mathbf{x})}$$

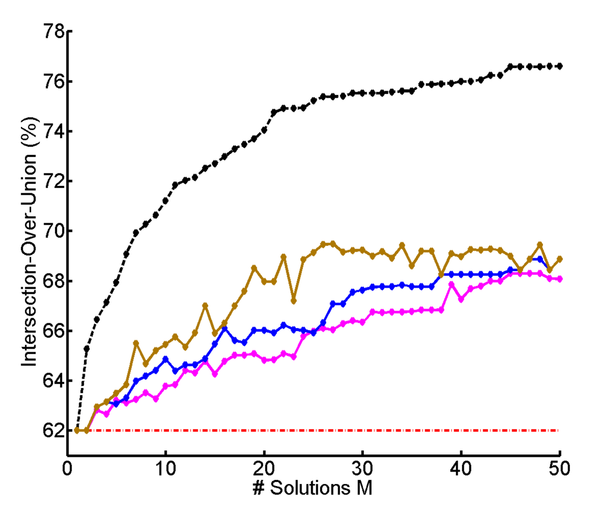
EMBR Predictor:

$$\mathbf{y}^{EMBR} = \arg\min_{\hat{\mathbf{y}} \in \mathbf{Y}} \sum_{\mathbf{y} \in \mathbf{Y}} \tilde{P}(\mathbf{y}|\mathbf{x})\tilde{\ell}(\mathbf{y}, \hat{\mathbf{y}})$$

Quantitative Results

Binary Segmentation

Dataset: Pascal VOC 2010 (100 images)
Model: 2-label Pairwise CRF
Inference: Graph Cuts

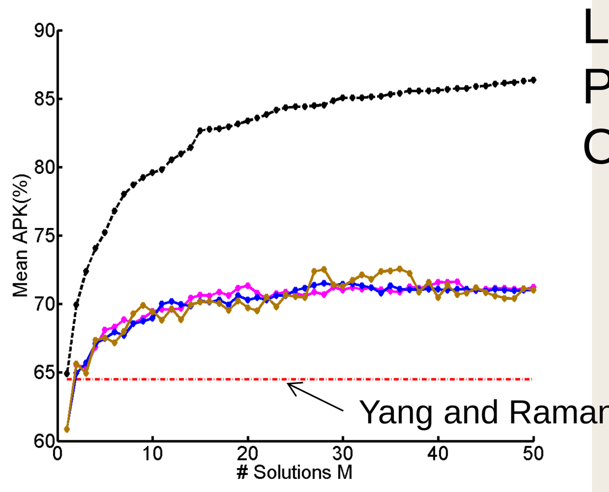


Loss Function: Intersection Over Union

~7% Gain

Pose Estimation

Dataset: PARSE Dataset
Model: Tree-structured parts model
Inference: Message passing

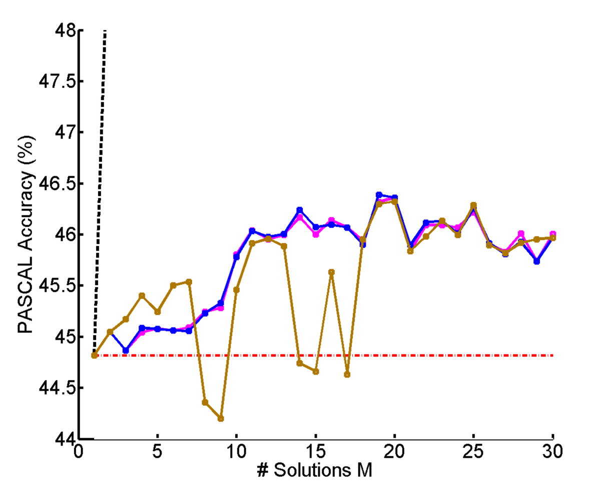


Loss Function: Percentage of Correct Parts

~7% Gain

Semantic Segmentation

Dataset: Pascal VOC 2012 Val Set
Framework: CPMC+O2P
Inference: Greedy pasting



Loss Function: Jaccard Index

~1.5% Gain

Oracle — MAP — 1-Param — 2-Params (λ, T) — 2-Params (λ, T) – Sensitivity Analysis