Monday p.m.

Poster #32

Classifier Adaptation at Prediction Time

Amélie Royer[1], Christoph H. Lampert[2]

[1] ENS Rennes (Ecole Normale Supérieure de Rennes), Rennes, France [2] IST Austria (Institute of Science and Technology Austria), Klosterneuburg, Austria

Motivation

Image classification with semantic context

- Efficient state-of-the-art image classifiers
- Designed for and tested on i.i.d. datasets
- real-life $\Longrightarrow$ more information, semantic visual context

Objective: Adapt a pre-trained classifier on-the-fly at prediction time to a realistic image sequence

Contributions:

- **Classifier adaptation scheme for three different settings: online, reinforced, unsupervised**
- **Dynamic adaptation for time-varying distributions**
- **Methods to generate sequences with realistic dependences**

Results: Online adaptation increases accuracy

Background

Domain Adaptation Problem

- Label distribution at training time, $P(y)$
- Label distribution at prediction time, $Q(y) \neq P(y)$
- Objects' visual appearance unchanged: $Q(x|y) = P(x|y)$

Framework

- $f : \mathcal{X} \to \mathcal{Y}$, pretrained multiclass probabilistic classifier

$$f(x) = \mathrm{argmax}_{y \in \mathcal{Y}} f_y(x) \quad \text{where } f_y \text{ reflects } P(\cdot\,|y)$$

Task: Adapt f to unknown label distribution $Q(y)$ on-the-fly

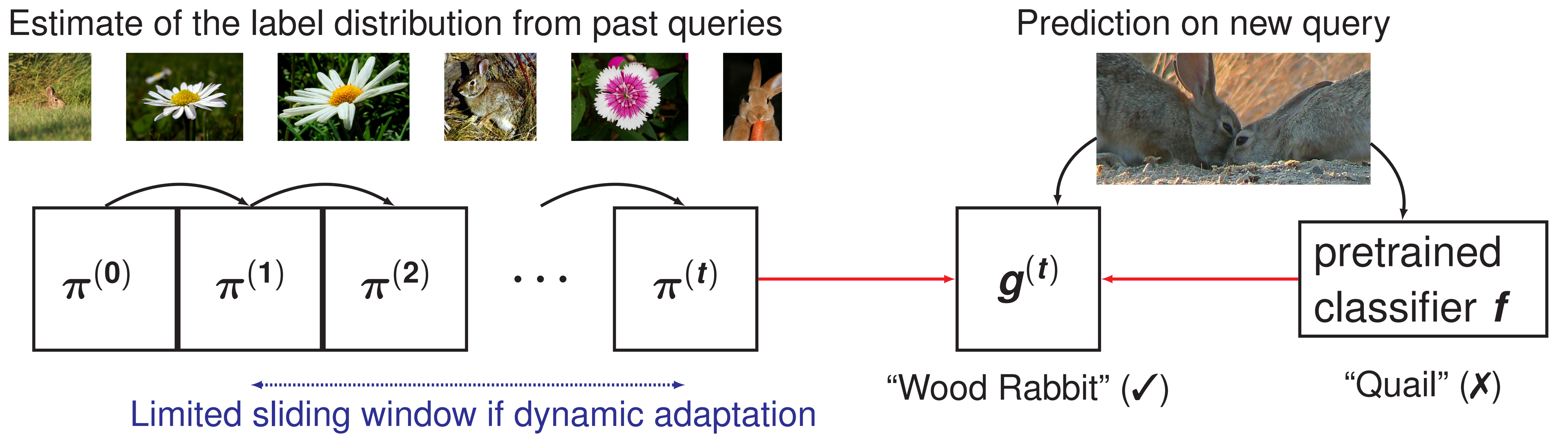
Related Work

- Identify relevant subset of class hierarchy on-the-fly [1]
- Adjusting classifier outputs to reflect new class priors [2]

[1] Y. Jia and T. Darrell. *"Latent task adaptation with large-scale hierarchies"*. ICCV, 2013.

[2] M. Saerens, P. Latinne, C. Decaestecker. *"Adjusting the outputs of a classifier to new a priori probabilities: a simple procedure"*, Neural Computation, 2002.

Workflow Example: Prediction a time t



Classifier Adaptation at Prediction Time

Class-prior adaptation

Definition. g is the Bayes-optimal class-prior adaptation of f at prediction time

$$g(x) = \mathrm{argmax}_{y \in \mathcal{Y}} g_y(x) \qquad \text{with } g_y(x) = \frac{Q(y) f_y(x)}{P(y)} \tag{1}$$

π_y: *online* estimation of the class distribution $Q(y)$ (Bayesian approach + Dirichlet prior $\mathbf{Dir}(\alpha)$)

$n_i(y)$: number of occurrences of class y among the i first queries.

$$\pi_y^{(i)} = \frac{n_i(y) + \alpha}{i + \alpha|\mathcal{Y}|}, \qquad \text{with } n_0(y) = 0 \text{ and } n_{i+1}(y) = n_i(y) + \delta_i(y) \tag{2}$$

Three online update scenarios

$$\text{(Online)} \quad \delta_i(y) = [\![y = y_i]\!] = 1 \text{ if } y = y_i, \text{ otherwise } 0 \tag{3}$$

$$\text{(Reinforced)} \quad \delta_i(y) = \begin{cases} [\![y = y_i]\!], & \text{if } y_i = g^{(i)}(x_i) \\ [\![y \neq y_i]\!]/(|\mathcal{Y}| - 1), & \text{otherwise} \end{cases} \tag{4}$$

$$\text{(Unsupervised)} \quad \delta_i(y) = \mathbb{E}_{\bar{y} \sim Q_i(\bar{y}|x_i)}([\![y = \bar{y}]\!]) = \frac{g_y^{(i)}(x_i)}{\sum_{\bar{y}} g_{\bar{y}}^{(i)}(x_i)} \tag{5}$$

Dynamic Adaptation

- *time-varying* data-distribution $\Rightarrow$ no static label distribution to converge to.
- *sliding window* of limited size $L \Rightarrow$ *dynamic* adaptation to recent labels only

$$\pi_y^{(i)} = \frac{n_i(y) + \alpha}{\min(i, L) + \alpha|\mathcal{Y}|}, \qquad \text{with } n_i(y) = \sum_{t=i-L+1}^{i} \delta_t(y) \tag{6}$$

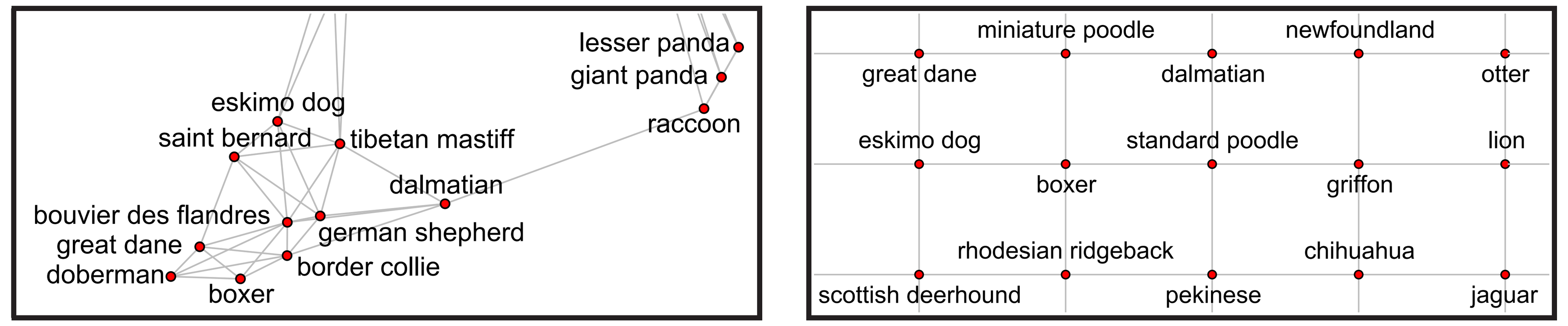
Generating Realistic Sequences

- **TXT**: Retrieve semantically ordered labels by browsing classical english books.

``*Alice* started to her *feet*, for it flashed across her *mind* that she had never before seen a <u>*rabbit*</u> with either a *waistcoat-pocket*, or a <u>*watch*</u> to take out of it''

"Alice in Wonderland" (L. Carroll) excerpt with ILSVRC2010 classes underlined

- **MDS and KS**: Random walk on a 2D projection (ImageNet's LCA distance). (MDS: *Multidimensional Scaling*, KS: projection with *Kernelized Sorting*)



Excerpts of the MDS (left) and KS (right) projections for ILSVRC2010 classes.

Experiments

- **Pretrained classifier**: CNN (CCV library) and SVM (JSGD toolkit)
- **Images** from ILSRVC2010 and 2012 validation datasets
- **Realistic Sequences**: 100 label sequences for TXT (average length 3475), KS, MDS (length 3000) **+** 100 random (RND) sequences.

		Online Feedback		Reinforcement		Unsupervised	
	CNN	Adapt	Dyn.	Adapt	Dyn.	Adapt	Dyn.
TXT	19.8 ± 1.9	$\mathbf{12.8 \pm 1.9}$	14.5 ± 1.7	$\mathbf{14.1 \pm 1.9}$	16.4 ± 1.7	$\mathbf{14.8 \pm 1.8}$	16.4 ± 1.7
MDS	16.1 ± 6.5	$\mathbf{5.2 \pm 3.0}$	6.6 ± 2.6	$\mathbf{6.1 \pm 3.4}$	8.3 ± 3.3	$\mathbf{6.8 \pm 3.6}$	8.1 ± 2.9
KS	16.4 ± 1.8	15.2 ± 1.7	$\mathbf{11.8 \pm 1.3}$	16.3 ± 1.8	$\mathbf{13.8 \pm 1.6}$	16.5 ± 1.7	$\mathbf{14.2 \pm 1.5}$
RND	$\mathbf{16.5 \pm 0.6}$	18.7 ± 0.7	17.2 ± 0.6	18.5 ± 0.7	16.8 ± 0.6	18.0 ± 0.7	16.9 ± 0.6

Top-5 error rates with the CNN classifier and ILSVRC2012. $\mathbf{10^{-3}}$*-significant entries marked in bold.*

Summary

Strengths

- simple and effective
- three different feedback scenarios
- works for arbitrary probabilistic classifiers

Limitations

- only marginal class probability
- what if new classes appear?
- guarantees only for *online* case