Minimax Forward and Backward Learning of Evolving Tasks with Performance Guarantees

Basque Center for Applied Mathematics-BCAM

Verónica Álvarez, valvarez@bcamath.org

Santiago Mazuelas, smazuelas@bcamath.org

Jose A. Lozano, jlozano@bcamath.org

Incremental learning a growing sequence of tasks

1910

Incremental learning a growing sequence of tasks

1910 **1940**

Incremental learning a growing sequence of tasks

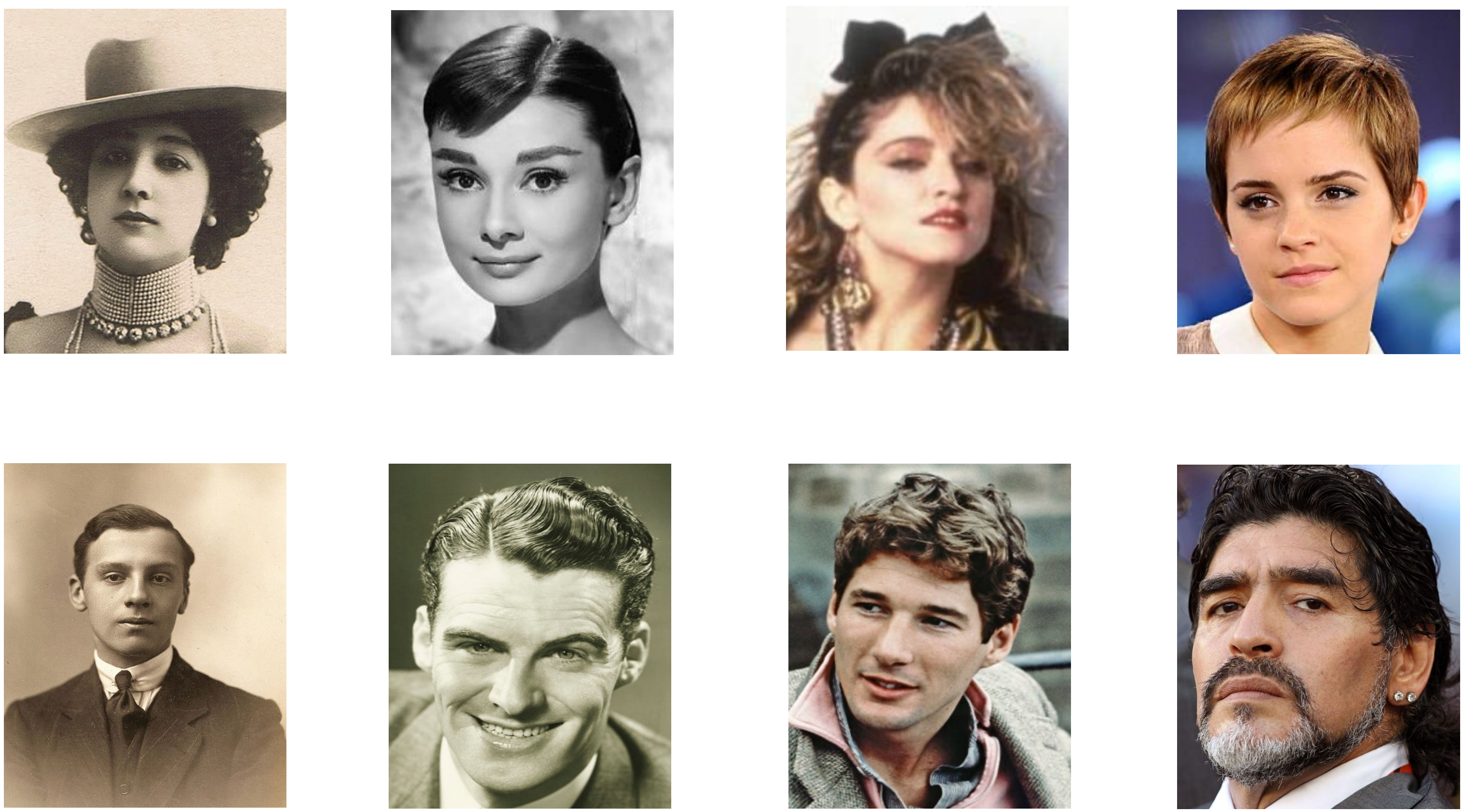
1910 **1940** **1980**





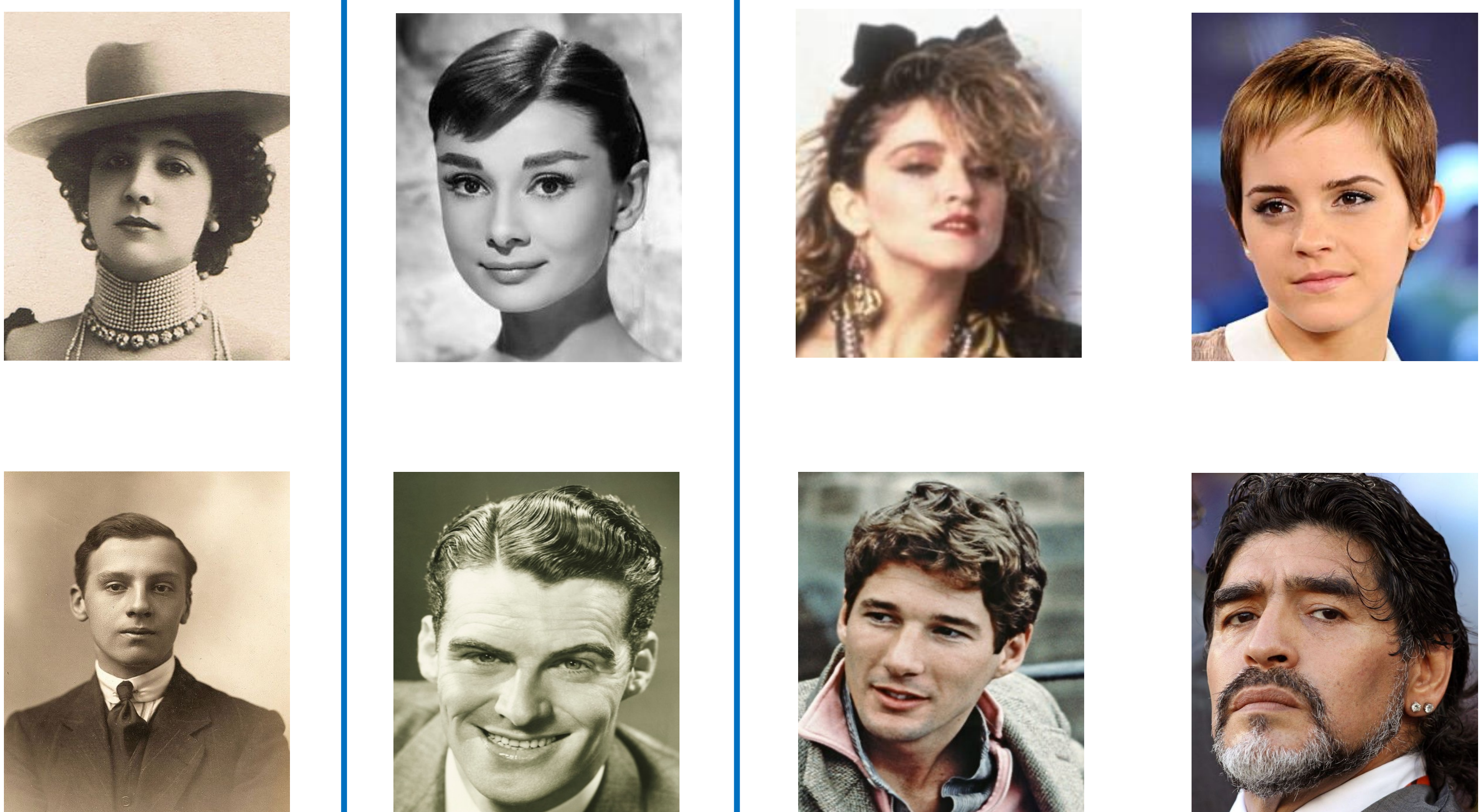

Incremental learning a growing sequence of tasks

1910 **1940** **1980** **2010**



Incremental learning a growing sequence of tasks

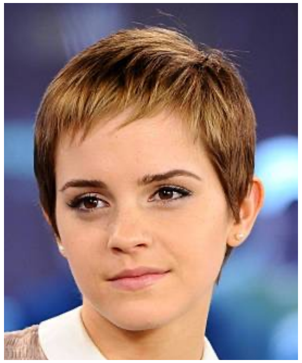
1910 **1940** **1980** **2010**



Incremental learning a growing sequence of tasks

1910 **1940** **1980** **2010**

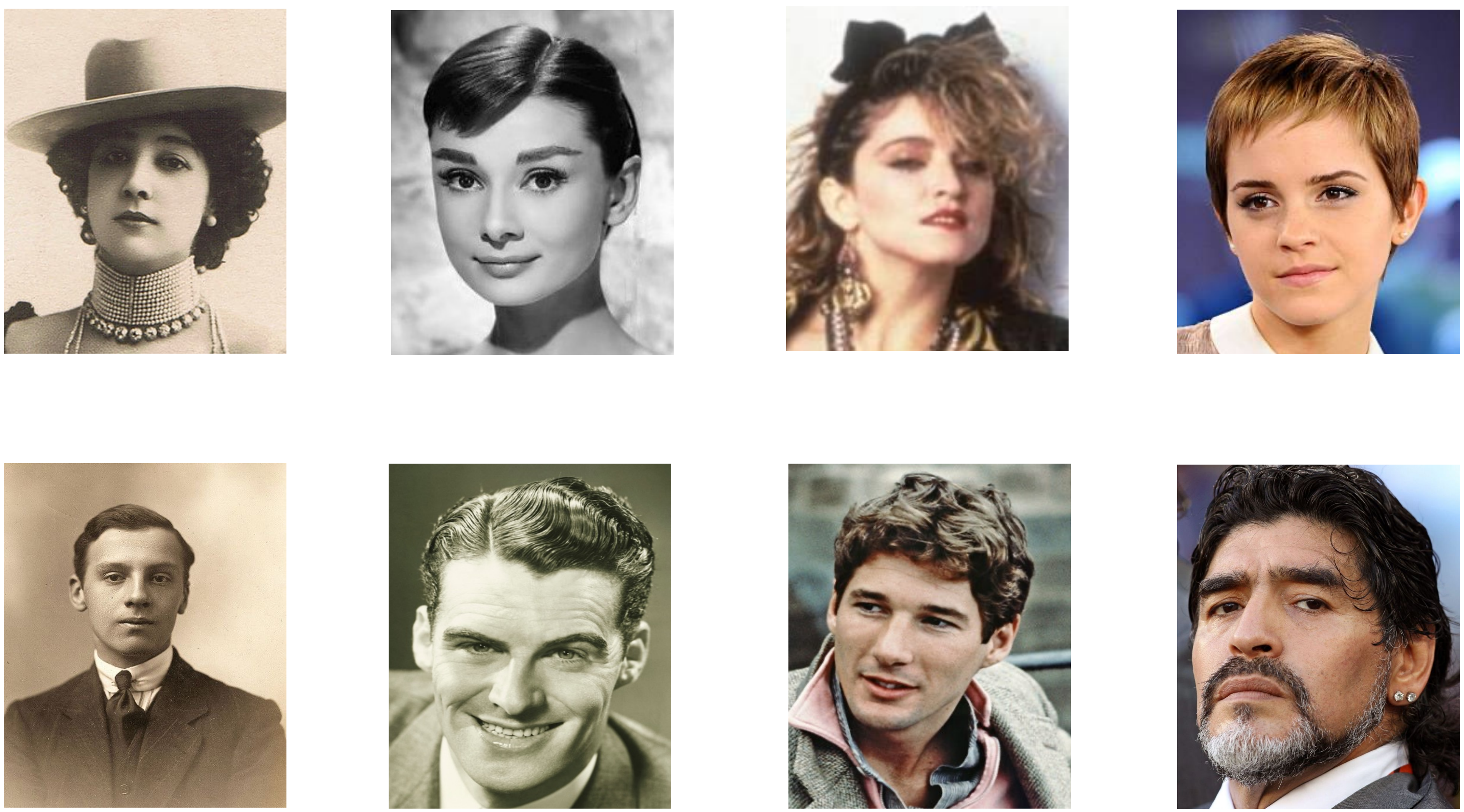





Incremental learning a growing sequence of tasks

1910 1940 1980 2010



Knowledge gaps

Concept drift adaptation techniques are designed for evolving tasks but only aim to learn the last task in the sequence

Continual learning techniques aim to learn the sequence of tasks but are not designed for evolving tasks

Knowledge gaps

Concept drift adaptation techniques are designed for evolving tasks but only aim to learn the last task in the sequence

Continual learning techniques aim to learn the sequence of tasks but are not designed for evolving tasks

Key contributions

Adapt to evolving tasks

Effectively exploit forward and backward learning

Provide performance guarantees and analytically characterize the increase in ESS

Incremental minimax risk classifiers (IMRCs)

Uncertainty set

$$\mathcal{U}_j^{\rightleftharpoons k} = \{\mathrm{p} \in \Delta(\mathcal{X} \times \mathcal{Y}) : |\mathbb{E}_{\mathrm{p}}\{\Phi(x, y)\} - \boldsymbol{\tau}_j^{\rightleftharpoons k}| \preceq \boldsymbol{\lambda}_j^{\rightleftharpoons k}\}$$

where $\Phi : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^m$ is a feature mapping

$\boldsymbol{\tau}_j^{\rightleftharpoons k}$ is the mean vector of expectation estimate

$\boldsymbol{\lambda}_j^{\rightleftharpoons k}$ is a confidence vector

Incremental minimax risk classifiers (IMRCs)

Uncertainty set

$$\mathcal{U}_j^{\rightleftharpoons k} = \{\mathrm{p} \in \Delta(\mathcal{X} \times \mathcal{Y}) : |\mathbb{E}_{\mathrm{p}}\{\Phi(x,y)\} - \boldsymbol{\tau}_j^{\rightleftharpoons k}| \preceq \boldsymbol{\lambda}_j^{\rightleftharpoons k}\}$$

where $\Phi : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^m$ is a feature mapping

$\boldsymbol{\tau}_j^{\rightleftharpoons k}$ is the mean vector of expectation estimate

$\boldsymbol{\lambda}_j^{\rightleftharpoons k}$ is a confidence vector

Learning

$$R(\mathcal{U}_j^{\rightleftharpoons k}) = \min_{\mathrm{h} \in \mathrm{T}(\mathcal{X},\mathcal{Y})} \max_{\mathrm{p} \in \mathcal{U}_j^{\rightleftharpoons k}} \ell(\mathrm{h}, \mathrm{p}) = \min_{\boldsymbol{\mu}} 1 - {\boldsymbol{\tau}_j^{\rightleftharpoons k}}^{\mathrm{T}} \boldsymbol{\mu} + \varphi(\boldsymbol{\mu}) + {\boldsymbol{\lambda}_j^{\rightleftharpoons k}}^{\mathrm{T}} |\boldsymbol{\mu}|$$

Prediction

$$\hat{y} \in \arg\max_{y \in \mathcal{Y}} \Phi(x,y)^{\mathrm{T}} \boldsymbol{\mu}_j^*$$

IMRCs

Single task learning

$$\boxed{\boldsymbol{\tau}_j} = \frac{1}{n_j} \sum_{i=1}^{n_j} \Phi(x_{j,i}, y_{j,i})$$

IMRCs

Single task learning

$$\boldsymbol{\tau}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\Phi(x_{j,i}, y_{j,i})$$

Forward learning

$$\boldsymbol{\tau}_{\overrightarrow{j}} = \boldsymbol{\tau}_j + \frac{\boldsymbol{s}_j}{\boldsymbol{s}_{\overrightarrow{j-1}} + \boldsymbol{s}_j + \boldsymbol{d}_j^2}\left(\boldsymbol{\tau}_{\overrightarrow{j-1}} - \boldsymbol{\tau}_j\right)$$

IMRCs

Single task learning

$$\boldsymbol{\tau}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\Phi(x_{j,i}, y_{j,i})$$

Forward learning

$$\boldsymbol{\tau}_j^{\rightharpoonup} = \boldsymbol{\tau}_j + \frac{\boldsymbol{s}_j}{\boldsymbol{s}_{j-1}^{\rightharpoonup} + \boldsymbol{s}_j + \boldsymbol{d}_j^2}(\boldsymbol{\tau}_{j-1}^{\rightharpoonup} - \boldsymbol{\tau}_j)$$

IMRCs

Single task learning

$$\boldsymbol{\tau}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\Phi(x_{j,i}, y_{j,i})$$

Forward learning

$$\boldsymbol{\tau}_{\overrightarrow{j}} = \boldsymbol{\tau}_j + \frac{\boldsymbol{s}_j}{\boldsymbol{s}_{\overrightarrow{j-1}} + \boldsymbol{s}_j + \boldsymbol{d}_j^2}\left(\boldsymbol{\tau}_{\overrightarrow{j-1}} - \boldsymbol{\tau}_j\right)$$

Forward and Backward learning

$$\boldsymbol{\tau}_{\overrightarrow{\overleftarrow{j}}^{k}} = \boldsymbol{\tau}_{\overrightarrow{j}} + \frac{\boldsymbol{s}_{\overrightarrow{j}}}{\boldsymbol{s}_{\overrightarrow{j}} + \boldsymbol{s}_{\overleftarrow{j+1}^{k}} + \boldsymbol{d}_{j+1}^2}\left(\boldsymbol{\tau}_{\overleftarrow{j+1}^{k}} - \boldsymbol{\tau}_{\overrightarrow{j}}\right)$$

Performance guarantees and effective sample size (ESS)

$$R(\mathcal{U}_j^{\rightleftharpoons k}) \leq R_j^{\infty} + \frac{M(\kappa+1)\sqrt{2\log\left(2m/\delta\right)}}{\sqrt{\mathrm{ESS}}} \left\|\boldsymbol{\mu}_j^{\infty}\right\|_1$$

Performance guarantees and effective sample size (ESS)

$$R(\mathcal{U}_j^{\rightleftharpoons k}) \leq R_j^{\infty} + \frac{M(\kappa+1)\sqrt{2\log(2m/\delta)}}{\sqrt{\text{ESS}}} \left\|\boldsymbol{\mu}_j^{\infty}\right\|_1$$

Single task learning n_j

Performance guarantees and effective sample size (ESS)

$$R(\mathcal{U}_j^{\rightleftharpoons k}) \leq R_j^{\infty} + \frac{M(\kappa+1)\sqrt{2\log(2m/\delta)}}{\sqrt{\text{ESS}}} \|\boldsymbol{\mu}_j^{\infty}\|_1$$

Single task learning n_j

Forward learning $n_j^{\rightharpoonup} \geq n_j + n_{j-1}^{\rightharpoonup} \frac{\|\boldsymbol{\sigma}_j^2\|_\infty}{\|\boldsymbol{\sigma}_j^2\|_\infty + n_{j-1}^{\rightharpoonup}\|\boldsymbol{d}_{j+1}^2\|_\infty}$

Performance guarantees and effective sample size (ESS)

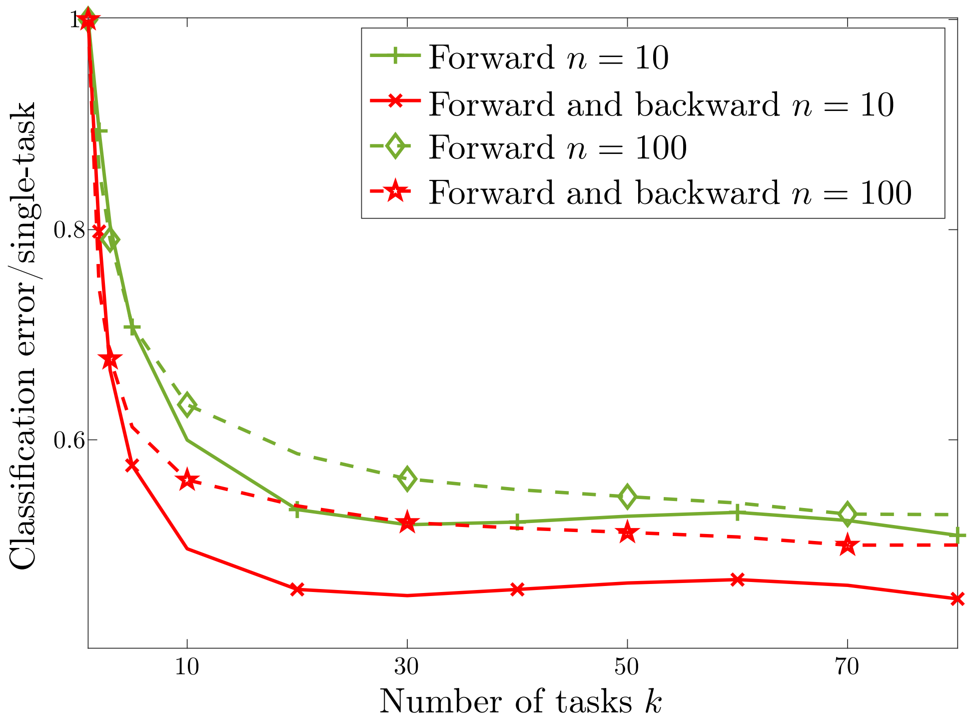
$$R(\mathcal{U}_j^{\rightleftharpoons k}) \leq R_j^{\infty} + \frac{M(\kappa+1)\sqrt{2\log\left(2m/\delta\right)}}{\sqrt{\mathrm{ESS}}} \left\|\boldsymbol{\mu}_j^{\infty}\right\|_1$$

Single task learning $\quad n_j$

Forward learning $\quad n_j^{\rightharpoonup} \geq n_j + n_{j-1}^{\rightharpoonup} \dfrac{\|\boldsymbol{\sigma}_j^2\|_\infty}{\|\boldsymbol{\sigma}_j^2\|_\infty + n_{j-1}^{\rightharpoonup}\|\boldsymbol{d}_{j+1}^2\|_\infty}$

Forward and Backward learning $\quad n_j^{\rightleftharpoons k} \geq n_j^{\rightharpoonup} + n_{j+1}^{\leftharpoondown k} \dfrac{\|\boldsymbol{\sigma}_j^2\|_\infty}{\|\boldsymbol{\sigma}_j^2\|_\infty + n_{j+1}^{\leftharpoondown k}\|\boldsymbol{d}_{j+1}^2\|_\infty}$

Experimental results



Algorithm	Yearbook	I. Noise	DomainNet	UTKFaces	R. MNIST	CLEAR
GEM	$.18 \pm .03$	$.39 \pm .08$	$.69 \pm .05$	$.12 \pm .00$	**$.36 \pm .06$**	$.57 \pm .10$
MER	$.16 \pm .03$	$.17 \pm .03$	$.38 \pm .04$	$.17 \pm .09$	$.37 \pm .09$	$.10 \pm .03$
ELLA	$.45 \pm .01$	$.48 \pm .05$	$.67 \pm .05$	$.19 \pm .12$	$.48 \pm .05$	$.61 \pm .03$
CL-MRC	**$.13 \pm .04$**	**$.15 \pm .03$**	**$.34 \pm .06$**	**$.10 \pm .01$**	**$.36 \pm .01$**	**$.09 \pm .03$**