Unsupervised Domain Adaptation for Personalized Facial Emotion Recognition

Gloria Zen[1], Enver Sangineto[1], Elisa Ricci[2], Nicu Sebe[1]

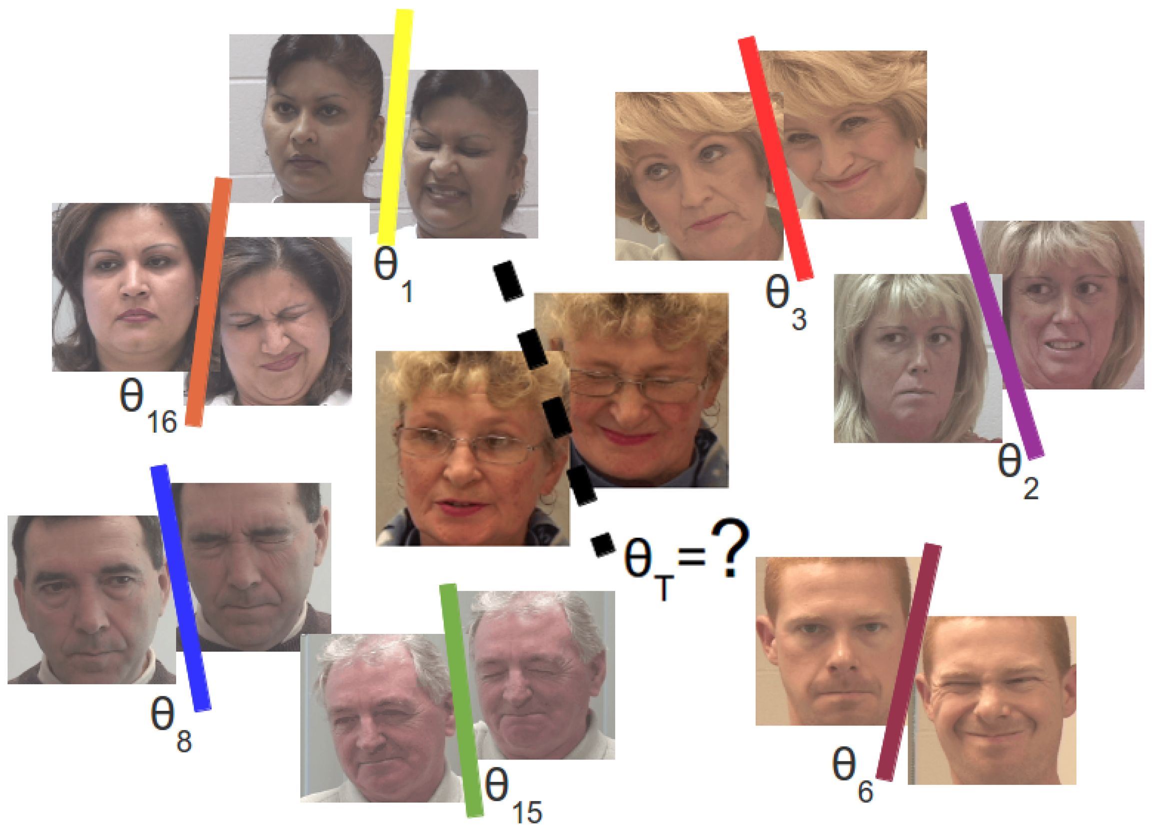
[1]{zen,sangineto,sebe}@disi.unitn.it [2]eliricci@fbk.eu

UNIVERSITÀ DEGLI STUDI DI TRENTO

FONDAZIONE BRUNO KESSLER

Motivations and Goals

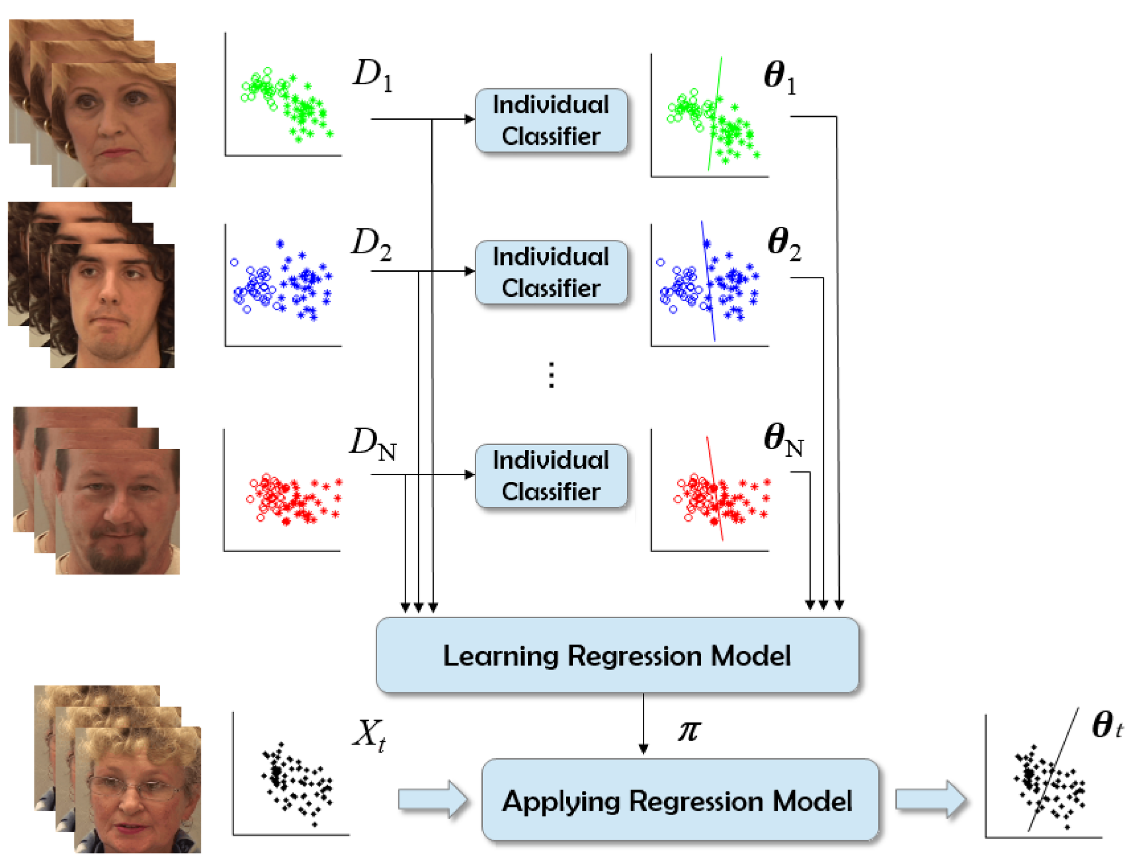
- Human expressions like pain or happiness can be exhibited differently, depending on the individual's appearance or personality. Previous works have shown that **person-specific** models are **advantageous** with respect to generic ones for recognizing facial expressions of new users added to the gallery set.
- Our goal is to obtain a personalized classifier $\boldsymbol{\theta}_t$ for a new user **without acquiring labeled data**.
- Unsupervised methods for domain adaptation usually rely on re-weighting source samples and retraining the classifier, which is a **time consuming** process. In some case scenarios, excessive waiting time may negatively affect **user experience quality**. We propose a personalization approach which is **drastically faster** w.r.t. related works.



- We show how $\boldsymbol{\theta}_t$ can be *accurately* and *efficiently* inferred **exploiting the similarity between** the data distribution of the target user and the **distributions** from other subjects with known $\boldsymbol{\theta}_i$.
- The intuition is that, despite the inter-subject variability, knowledge can be transferred among individuals showing similar behavioral patterns.

SV-based Transductive Parameter Transfer (SVTPT)

- Let $\mathcal{X}, \mathcal{Y}$ be, respectively, a feature and a label space, with $\mathcal{Y} = \{-1, 1\}$.
- We assume to have N *labeled source* datasets $D_1^s, ..., D_N^s$, $D_i^s = \{\mathbf{x}_j^s, y_j^s\}_{j=1}^{n_i^s}$, and an *unlabeled target* dataset $X^t = \{\mathbf{x}_j^t\}_{j=1}^{n^t}$
- We assume that the vectors in $\mathbf{x}_j^{\{s,t\}}$ are generated by a marginal distribution $P_i^{\{s,t\}}$ defined on $\mathcal{X}$.



The approach we propose is based **on three main steps**:

1. A set of source-specific classifiers $\boldsymbol{\theta}_i^s$ is learned for each user
2. A regression algorithm is adopted to learn the relation between the marginal distributions P_i^s and the source classifiers' parameter vectors $\boldsymbol{\theta}_i^s$.
3. The desired target classifier is obtained by applying the learned **distribution-to-classifier mapping function** and using as input the data points of the new target user.

Kernel Estimation

Density Estimate-based Kernel. A kernel measuring the similarity of two distributions is defined as follows:

$$\kappa_{DE}(X_i, X_j) = \frac{1}{nm} \sum_{p=1}^{n} \sum_{q=1}^{m} \kappa_{\mathcal{X}}(\mathbf{x}_p, \mathbf{x}_q),$$

Learning from Distributions

The **mapping function** $f : \mathcal{P} \to \Theta$ can be defined as a set of parameters $\boldsymbol{\pi}_k = (\mathbf{b}_k, c_k)$:

$$\widehat{f}_k(X) = \langle \phi_k(X), \mathbf{b}_k \rangle + c_k \tag{1}$$

where $\phi_k(X)$ is a nonlinear mapping of X to a higher-dimensional space.
In turn $\boldsymbol{\pi}_k$ can be found by minimizing:

$$\min_{\boldsymbol{\pi}} \frac{1}{2}||\boldsymbol{b}_k||^2 + \lambda_E \sum_{i=1}^{N} |\boldsymbol{\theta}_{ik} - \widehat{f}_{\boldsymbol{\pi}}(Z_i)|_\epsilon \tag{2}$$

Eq. (2) can be transformed into the dual problem:

$$\max_{\{\beta_i^k\}} -\frac{1}{2} \sum_{i,l=1}^{N} \beta_i^k \beta_l^k \kappa(Z_i, Z_l) + \sum_{i=1}^{N} \boldsymbol{\theta}_{ik} \beta_i^k - \epsilon \sum_{i=1}^{N} |\beta_i^k| \tag{3}$$

where $Z^t = X^t$ for the *new unseen* target users, and $Z_i^s = V_i = \{\mathbf{v}_j\}_{j=1}^{m_i}$, i.e. the Support Vectors associated with $\boldsymbol{\theta}_i$ ($V_i \subseteq X_i^s$).

Results

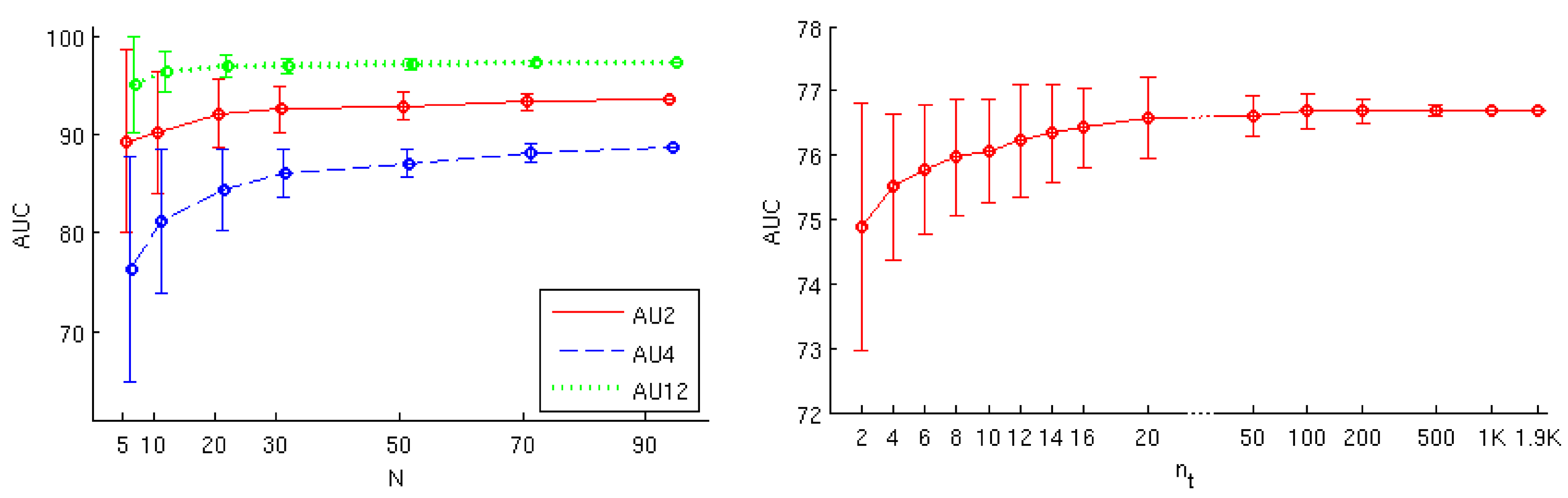
- Performances on Action Unit Detection (CK+ dataset) measured in AUC:

AU	SVM	KMM [1]	TSVM [2]	DASVM [3]	STM [4]	TPT [6]	SVTPT Allr	SVTPT SVs
1	79.8	68.9	69.9	72.6	88.9	88.2	87.8	89.6
2	90.8	73.5	69.3	71.0	87.5	92.6	92.8	93.9
4	74.8	62.2	63.4	79.9	81.1	84.3	84.5	88.6
6	89.7	87.7	61.5	94.7	94.0	91.7	91.1	91.5
12	88.1	89.5	76.0	95.5	92.8	97.1	96.3	97.5
17	90.3	66.6	73.1	94.7	96.0	94.3	94.4	94.1
Avg	85.6	74.7	68.7	83.1	90.1	91.3	91.3	**92.7**

- Performances on Pain Facial Expression Recognition (UNBC-MSPEAD dataset) measured in AUC:

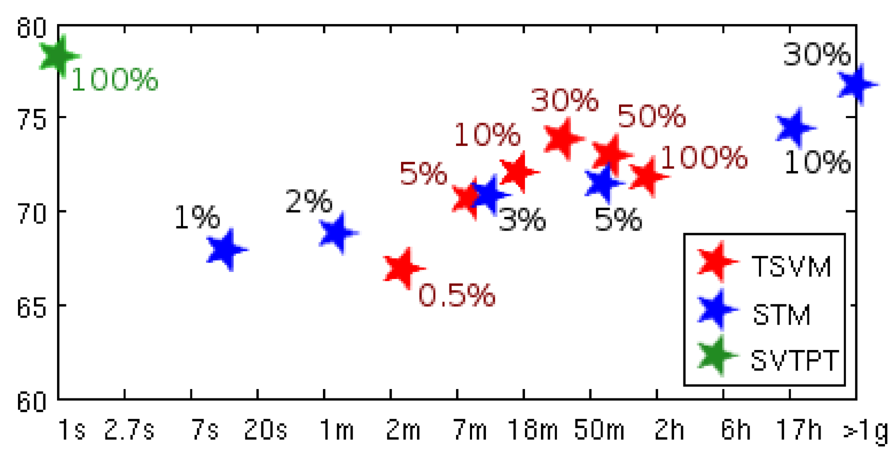
SVM	TTA [5]	TSVM [2]	STM [4]	TPT [6]	SVTPT All	SVTPT SVs
75.6	76.5	69.8	76.8	76.7	76.7	**78.4**

Performance at varying number of samples



(*Left*) Performance of SVTPT on CK+ dataset at varying number of **source users** N. (*Right*) Performance of SVTPT on UNBC-MSPEAD dataset at varying number of **target samples** n_t.

Performances and Computational Costs



UNBC-MSPEAD dataset. Performance (AUC) vs average time (in logarithmic scale) for training a target classifier with different unsupervised personalization methods.

References

[1] A. Gretton, et al. Dataset shift in machine learning In *MIT press Cambridge*,, 2009.

[2] J. Chen, X. Liu and P. Tu, A. Aragones. Learning person-specific models for facial expression and action unit recognition In *Pattern Recognition Letters*, 2013.

[3] L. Bruzzone, M. Marconcini Domain adaptation problems: A DASVM classification technique and a circular validation strategy In *PAMI*, 2010

[4] W.S. Chu, F. De La Torre, J.F. Cohn. Selective transfer machine for personalized facial action unit detection In *CVPR*, 2013.

[5] Joachims, Thursten Transductive inference for text classification using support vector machines In *ICML*, 1999.

[6] E. Sangineto, G. Zen, E. Ricci, N. Sebe. We are all Not Equal: Personalized Models for Facial Expression Analysis via Transductive Parameter Transfer In *ACM MM (long paper)*, 2014.