We are not All Equal: Personalizing Models for Facial Expression Analysis with Transductive Parameter Transfer

Enver Sangineto[1], **Gloria Zen**[1], Elisa Ricci[2,3], Nicu Sebe[1]

UNIVERSITY OF TRENTO

[1] DISI, University of Trento, Italy
[2] DIEI, University of Perugia, Italy
[3] Fondazione Bruno Kessler (FBK), Italy

FONDAZIONE BRUNO KESSLER

Motivations and goals

- **Human expressions** (e.g., pain) are exhibited diversely, **depending on** the **individual's appearance** or **personality.** Previous work has proven that **personalized classifiers** perform **better than generic ones**

- Our ***goal*** is to obtain a personalized classifier for a new user **without acquiring new labeled data**

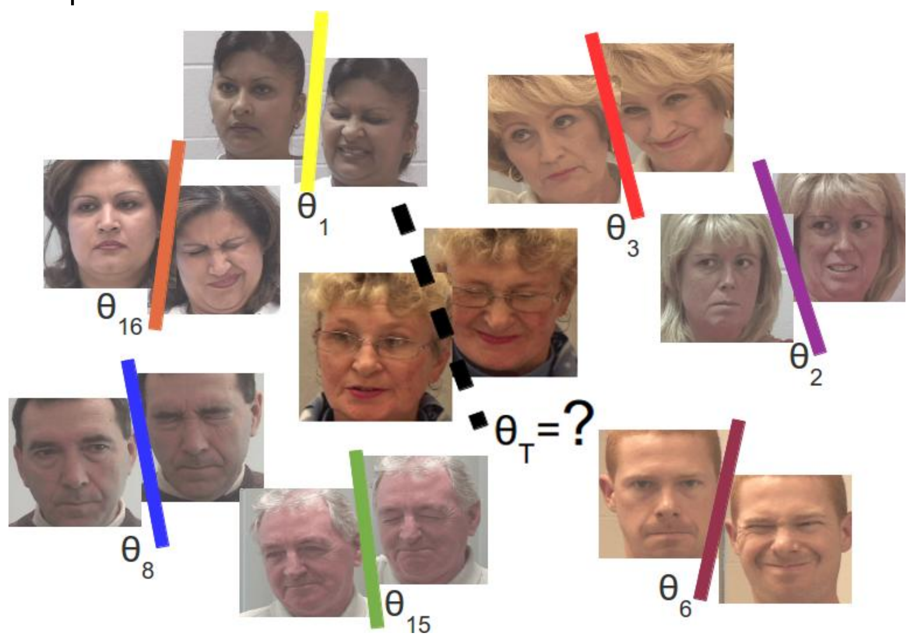
- Most of related work rely on re-weighting source samples and retraining a classifier, which is a **time consuming** process. A faster personalization is needed in some cases, e.g. where timing can affect user experience quality.

Sample frames of facial (top) non-pain and (bottom) pain expressions.

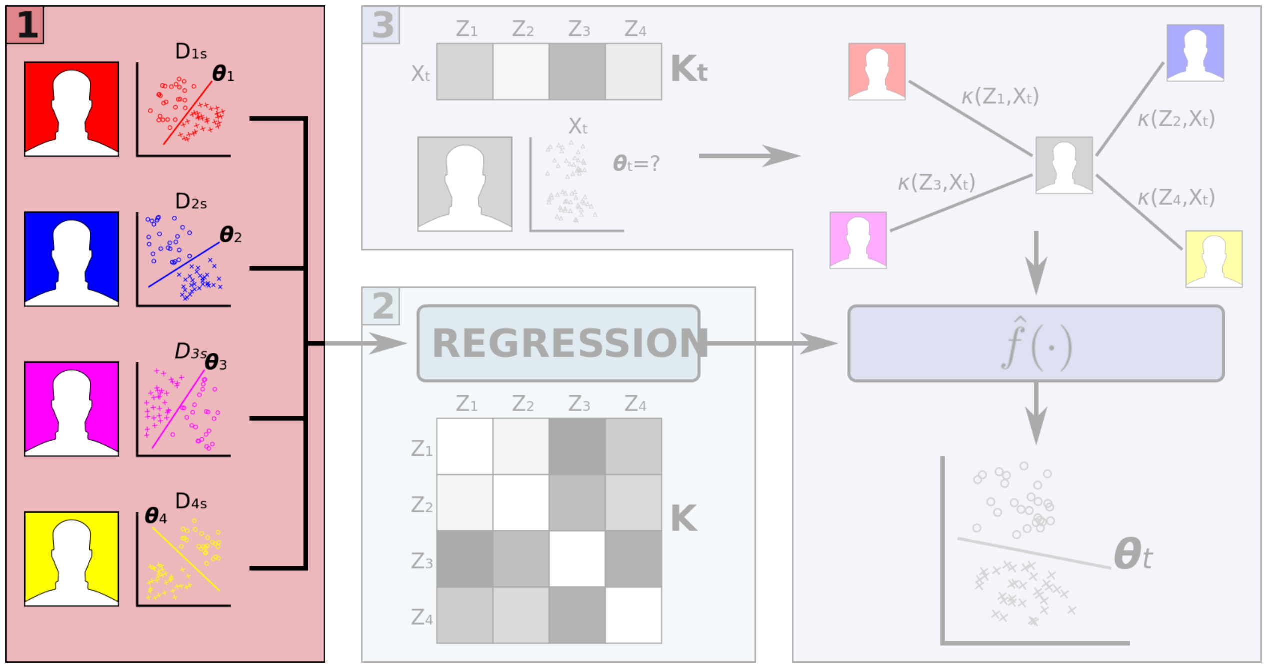
Intuition behind our approach

- Despite the inter-subject variability, **knowledge can be transferred among individuals** showing similar behavioral patterns.

- We show how θ_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 θ_i

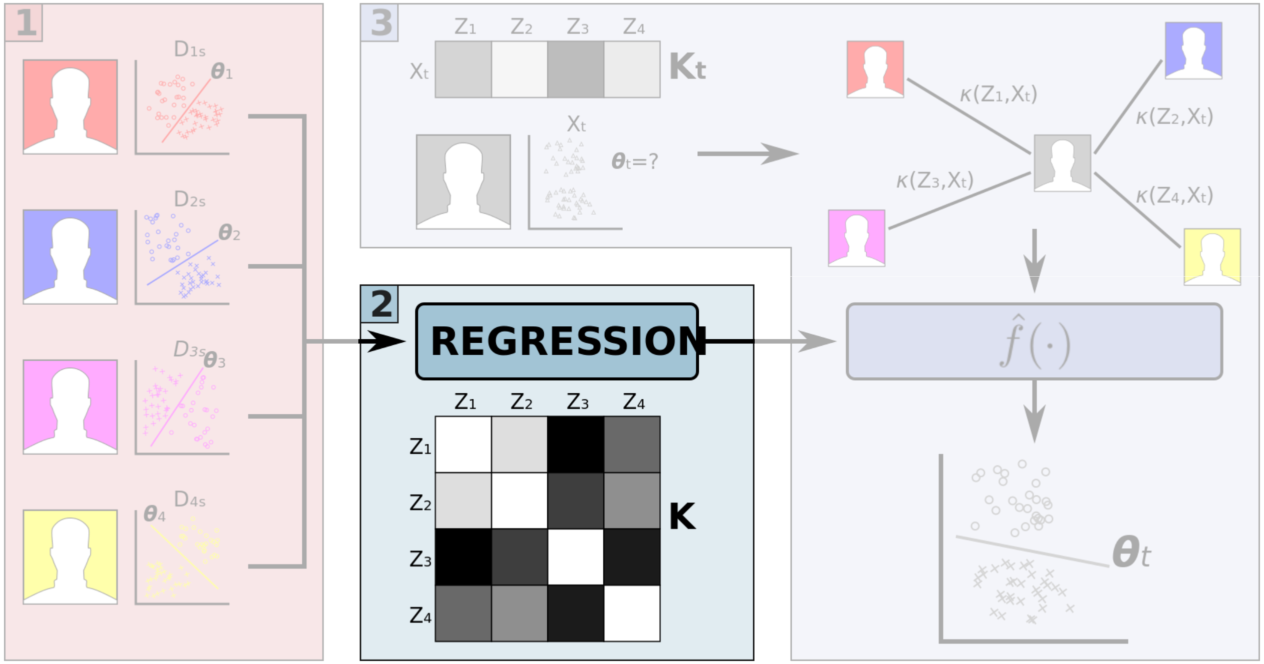


Transductive Parameter Transfer (TPT) 1/3



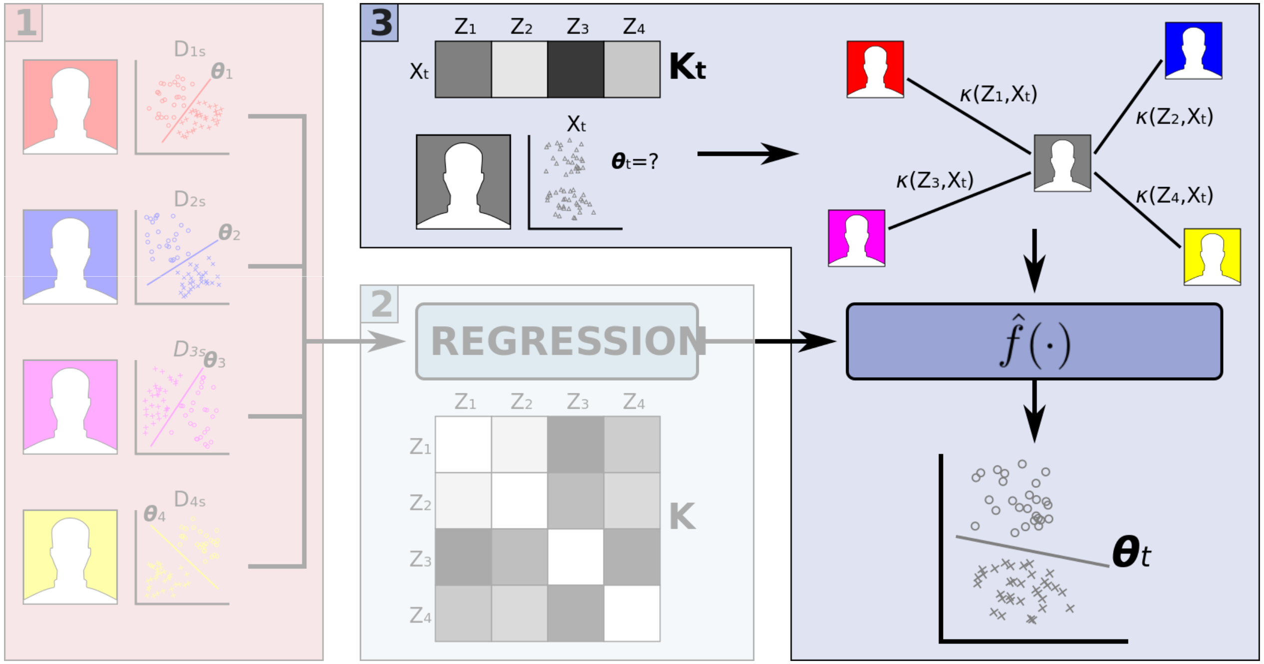
(1) A **set of personalized** facial expression **classifiers** is learnt for each **source user**.

Transductive Parameter Transfer (TPT) 2/3



(**2**) We learn via a regression framework a **mapping between** a **marginal distribution** of the datapoints associated to a given person and the **parameters** of her/his **personalized classifier**.

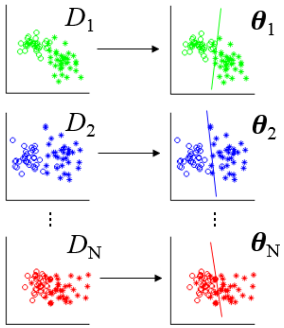
Transductive Parameter Transfer (TPT) 3/3



(**3**) A **personalized classifier** is computed for a **new given target** by **applying** the learnt **distribution-to-classifier mapping**.
A **kernel function** measures the **users' similarity** based on their data points distributions.

Learning from Distributions 1/2

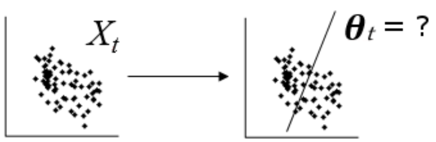
❖ Input data



- $\boldsymbol{N}$ labelled **source datasets** with distributions $\boldsymbol{D}_\mathrm{i} = \{\boldsymbol{X}_\mathrm{i}, \boldsymbol{Y}_\mathrm{i}\}$ and learnt personalized classifiers $\boldsymbol{\Theta}_\mathrm{i} = \{\mathbf{w}_\mathrm{i}, \boldsymbol{b}_\mathrm{i}\}$

❖ Learning a **distribution-to-classifier mapping** $f : \mathcal{P} \rightarrow \Theta$

- $\mathcal{P}$ is the **space** of all the possible **distributions** on $\boldsymbol{X}$
- Θ is the **classifier parameter space**
- Learning a relationship between the "shape" of the underlying distribution and its corresponding hyperplane.
- Once $f(\cdot)$ is estimated, it can be applied on the new target data distribution for the hyperplane estimation



Learning from Distributions 2/2

- We use Multioutput Support Vector Regression (M-SVR)[1] for the mapping function estimation. The **mapping function** $f : \mathcal{P} \rightarrow \Theta$ is defined by a set of parameters:

$$\widehat{f}(X) = \phi(X)'\mathbf{B} + \mathbf{c}' \quad ; \quad \text{where} \begin{cases} \boldsymbol{\pi} = (\mathbf{B}, \mathbf{c}) \\ \phi(X) \ \text{is a non-linear mapping} \end{cases}$$

- Parameters can be found by **minimizing**:

$$\min_{\boldsymbol{\pi}} \frac{1}{2} \sum_{i=1}^{M+1} ||\boldsymbol{\beta}_i||^2 + \lambda_E \sum_{i=1}^{N} E(\|\boldsymbol{\theta}'_i - \widehat{f}_{\boldsymbol{\pi}}(X_i)\|)$$

- The same problem can be solved by introducing **kernel matrix** K:

$$\widehat{f}(X) = \sum_{i=1}^{N} \mathbf{V}_{i.}\kappa(X_i, X) + \mathbf{c}'$$

where $\mathbf{K}_{ij} = \kappa(X_i, X_j) = \phi(X_i)'\phi(X_j); \quad \mathbf{K} \in I\!R^{N \times N}$

[1] D. Tuia, et al. *Multioutput support vector regression for remote sensing biophysical parameter estimation*. IEEE Geoscience and Remote Sensing Letters, 8(4):804–808, July 2011.

Kernel Matrix Estimation

❖ Representing similarity between pairs of datasets $K(\boldsymbol{X}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{j}})$

- **Fisher** Kernel

$$\kappa_{FK}(X_i, X_j) = (\mathcal{G}_\gamma^{X_i})' \mathcal{G}_\gamma^{X_j}$$

- **Earth Mover's Distance (EMD)** based Kernel

$$\kappa_{EMD}(X_i, X_j) = e^{-\rho D_{EMD}(X_i, X_j)}$$

- **Density Estimate (DE)** based Kernel

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

Application Domains and Datasets

- ## Facial Action Unit Detection

Extended Cohn Kanade (CK+)

- set of spontaneous and posed expressions
- 593 videos, 123 users
- Facial landmark detection + SIFT features

- ## Pain Facial Expression Recognition

UNBC-McMaster Shoulder Pain Expression Archive Database (UNBC-MSPEAD)

- spontaneous pain expressions of patients under shoulder mobility tests
- 200 videos, 25 users
- Local Binary Patterns Histograms features

Results – Action Unit Detection

F-Score

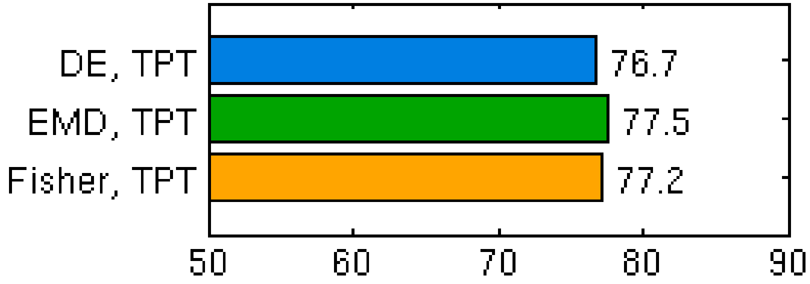
AU	SVM	KMM	TSVM	DASVM	STM	TPT EMD	TPT Fisher	TPT DE
1	61.1	44.9	56.8	57.7	62.2	72.2	74.0	74.4
2	73.5	50.8	59.8	64.3	76.2	81.8	75.5	84.2
4	62.7	52.3	51.9	57.7	69.1	71.5	71.8	66.3
6	75.7	70.1	47.8	68.2	79.6	75.1	74.9	74.8
12	76.7	74.5	59.6	59.0	77.2	85.5	83.5	85.1
17	76.0	53.2	61.7	81.4	84.3	82.8	83.5	76.1
Avg	70.9	57.6	56.3	64.7	74.8	**78.2**	77.2	76.8

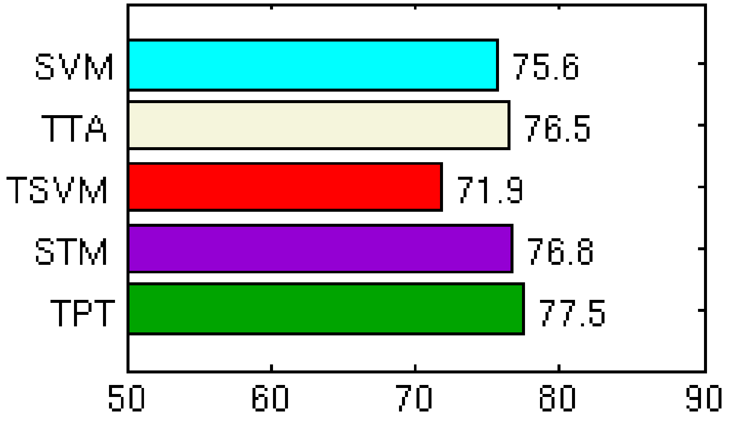
AUC

AU	SVM	KMM	TSVM	DASVM	STM	TPT EMD	TPT Fisher	TPT DE
1	79.8	68.9	69.9	72.6	88.9	88.0	89.0	88.2
2	90.8	73.5	69.3	71.0	87.5	93.5	92.9	92.6
4	74.8	62.2	63.4	79.9	81.1	88.1	85.0	84.3
6	89.7	87.7	60.5	94.7	94.0	92.2	91.3	91.1
12	88.1	89.5	76.0	95.5	92.8	97.5	97.2	97.1
17	90.3	66.6	73.1	94.7	96.0	95.9	94.3	94.3
Avg	85.6	74.7	68.7	83.1	90.1	**92.5**	91.6	91.3

[KMM] Gretton et al. “*Covariate shift by kernel mean matching*”. Dataset shift in Machine Learning, 2009
[TSVM] T. Joachim “*Transductive inference for text classification using support vector machines*”, ICML, 1999
[DASVM] Bruzzone et al. “*Domain adaptation problems: A DASVM classification technique and a circular validation strategy*”. TPAMI, 2010.
[STM] W.S. Chu et al. “*Selective transfer machine for personalized facial action unit detection*”. CVPR, 2013.

Results – Pain Facial Expression Recognition

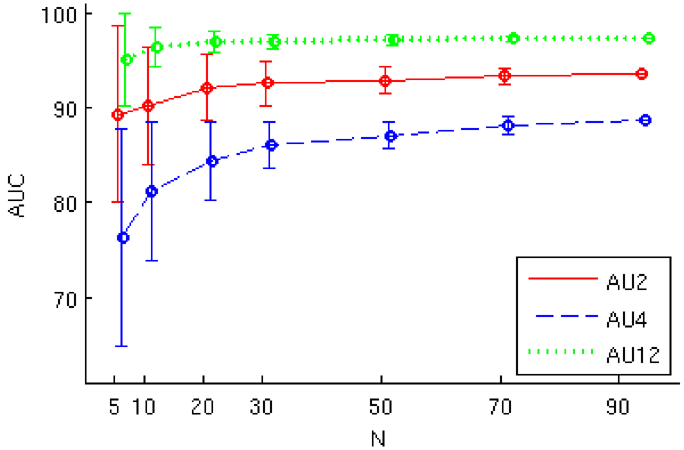




[TTA] J. Chen et al. “*Learning person-specific models for facial expression and action unit recognition*”. Pattern Recognition Letters, 2013
[TSVM] T. Joachim “*Transductive inference for text classification using support vector machines*”, ICML, 1999
[STM] W.S. Chu et al. “*Selective transfer machine for personalized facial action unit detection*”. CVPR, 2013.

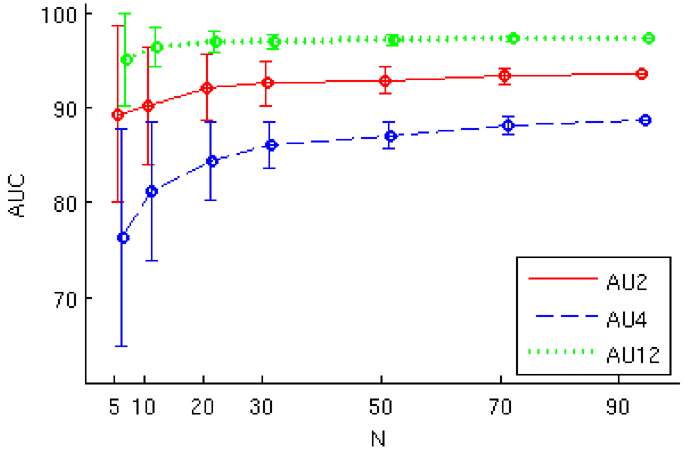
Results at varying number of source data

- Performances at varying number of users N (CK+ dataset)

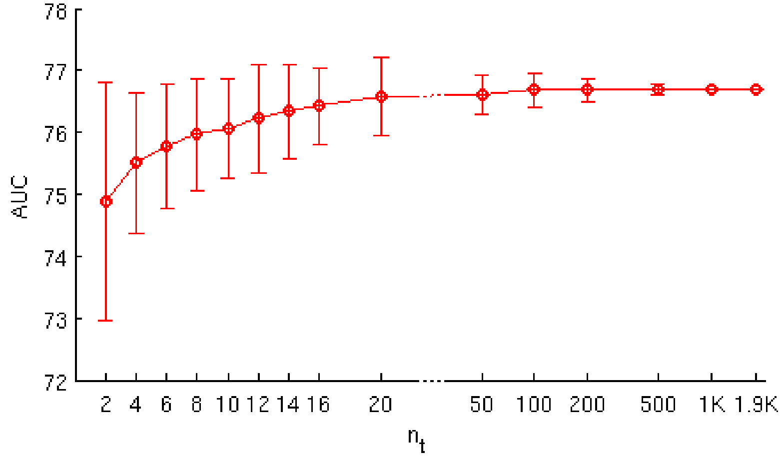


Results at varying number of source data

- Performances at varying number of users N (CK+ dataset)

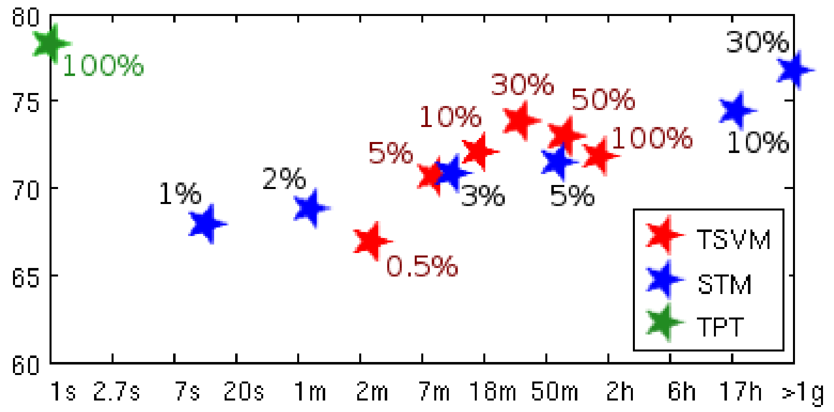


- Performances at varying number of training images for a new target user (UNBC-MSPEAD dataset)



Performance and Computational Cost

- Performances vs computational time. Comparison among related works



Our method is more efficient because it is not based on retraining the classifier!

[TSVM] T. Joachim "*Transductive inference for text classification using support vector machines*", ICML, 1999
[STM] W.S. Chu et al. "*Selective transfer machine for personalized facial action unit detection*". CVPR, 2013.

Conclusions

- we proposed a novel domain adaptation approach for facial expression analysis which deals with the inter-individual variability
- A classifier for a new target individual is inferred without the need of acquiring labeled data
- our system achieves state-of-the-art accuracy on public benchmarks while being different orders of magnitude faster than other unsupervised domain adaptation approaches

Q & A

THANK YOU
FOR
your
ATTENTION
ANY QUESTIONS