PARTIAL LEAST SQUARES BASED RANKER FOR FAST AND ACCURATE AGE ESTIMATION

Hong Liu, Xiaohu Sun

Key Laboratory of Machine Perception, Shenzhen Graduate School, Peking University, China

{hongliu, xiaohusun}@pku.edu.cn **Presenter**: Xiaohu Sun

• INTRODUCTION

1

- Facial age estimation
- Useful for friendly and secure human robot / computer interactions, age-based access control, family photo management, etc.
- Challenging due to complex dynamics in aging process.

• MOTIVATION

- PLS has many excellent characteristics.
- Rankers show favorable performance by exploiting the ordinal nature of ages.
- Transform the advanced multivariate data analysis (MVA) tool PLS, into a ranker to boost the performance in terms of accuracy, speed and robustness, on the age estimation problem.

• Advantages of PLS based ranker (PLS-Ranker)

- **All binary classifiers are jointly learned.**
- **The proposed adaptive threshold learning strategy makes each basic binary classifier of PLS-Ranker robust to the imbalance problem of its positive and negative training data. Thus boosts ranking accuracy of PLS-Ranker.**
- **PLS-Ranker simultaneously reduces feature dimension and ranks in high speed even for high-dimensional features.**

PARTIAL LEAST SQUARES BASED RANKER

2

- Feature matrix $\mathbf{X}$ $(n \times N)$: feature vectors of n training samples.
- Column vector $\boldsymbol{y}$ $(n \times 1)$: scalar age values of n training samples.
- Suppose the age range is $1 \sim \mathrm{M}_a$.

1. Encoding age for jointly learning all associated binary classifiers

- Encode each scalar age y_i $(i = 1, 2, \cdots, n)$ into a $1 \times (\mathrm{M}_a - 1)$ binary row vector $\mathbf{A}(i,:)$.
- Form an $n \times (\mathrm{M}_a - 1)$ indicator matrix $\mathbf{A}$.
- $\mathrm{A}_{i,j}$ indicates whether the face i is older than j years.

$$\mathrm{A}_{i,j} = \begin{cases} 1, & \text{if } j < y_i \\ 0, & \text{if } j \geq y_i \end{cases}$$

2. Train a linear PLS model

- Use the feature matrix $\mathbf{X}$ and indicator matrix $\mathbf{A}$ to train a linear PLS model .
- Get the regression coefficient matrix $\mathbf{B}_{pls}$.

3. Adaptive threshold learning

• Imbalance problem

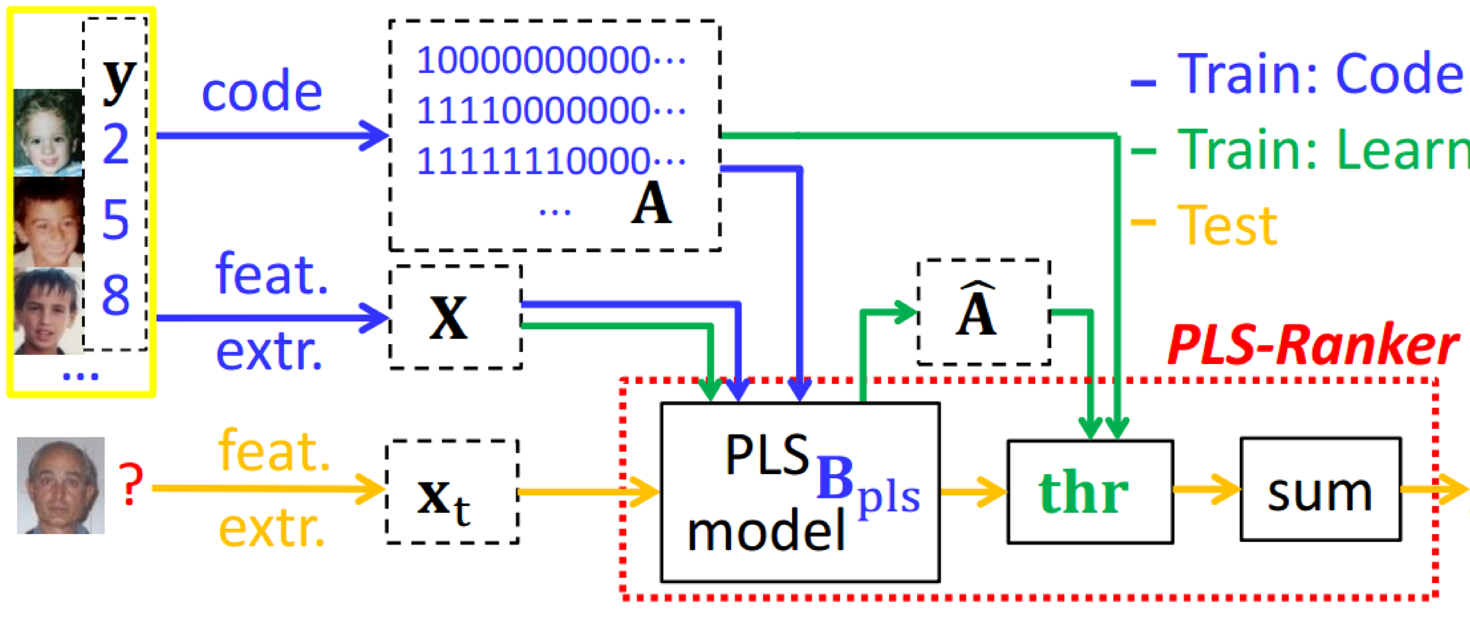
- Reduce ranking to associated binary classifications: Is the face older than *j years*? $j = 1, 2, \cdots, \mathrm{M}_a - 1$
- Under the reduction framework, the key to improve ordinal ranking is to improve binary classifications.
- However, the positive and negative training samples for each of the binary classifiers can be highly unbalanced.
- The unbalanced positive and negative training samples may shift the optimal thresholds away from 0.5.

• Learn thresholds from the unbalanced training data

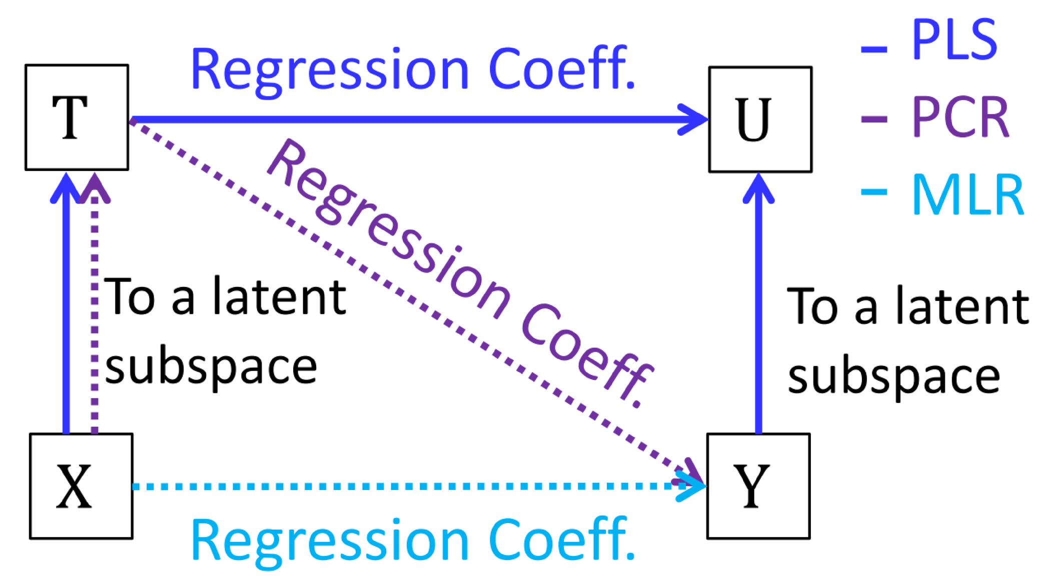
- Apply the trained linear PLS model back to the feature matrix of training set $\mathbf{X}$
- Get a prediction of indicator matrix $\mathbf{A}$, $\hat{\mathbf{A}}$:

$$\hat{\mathbf{A}} = \mathbf{X}\mathbf{B}_{pls}$$

- $\hat{\mathbf{A}}$ is a real-valued matrix.
- Thresholds are searched in a small range S around 0.5.
- Determined by the criterion of minimizing the error rate on the training set.



(a) Flow chart of age estimation using PLS-Ranker



(b) MLR, PCR and PLS coefficients calculation

3

For the j-th $(j = 1, 2, \cdots, M_a - 1)$ binary classifier:

$$\mathrm{thr}_j = arg \max_{b \in S} \sum_{i=1}^{n} f_m(\mathrm{A}_{i,j}) \times f_m([\![\hat{\mathrm{A}}_{i,j} \geq b]\!])$$

• Test process

- $\mathbf{x}_t$ $(1 \times N)$ is mapped to an indicator vector $\hat{\mathbf{a}}_t$ using the trained linear PLS model:

$$\hat{\mathbf{a}}_t = \mathbf{x}_t \mathbf{B}_{pls}$$

- Threshold each element of $\hat{\mathbf{a}}_t$ into a binary indicator:

$$\hat{\mathbf{a}}_t(1,j) \coloneqq [\![\hat{\mathbf{a}}_t(1,j) \geq \mathrm{thr}_j]\!] \qquad j = 1, 2, \cdots, \mathrm{M}_a - 1$$

- The rank (age) is obtained by summarizing elements in $\hat{\mathbf{a}}_t$: $\hat{r}_t = \sum_{j=1}^{\mathrm{M}_a - 1} \hat{\mathbf{a}}_t(1,j) + 1$

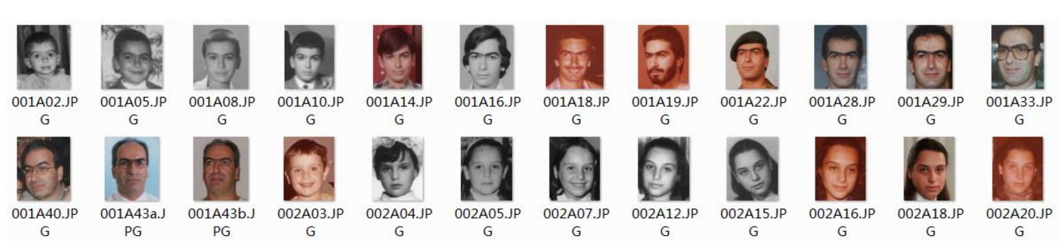
EXPERIMENTS AND DISCUSSIONS

4

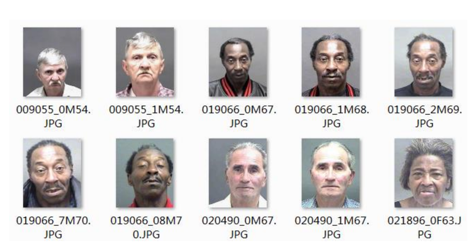
• Datasets

- **FG-NET**. Leave one person out.
- **MORPH**. **Setting 1** and **Setting 2.**

FG-NET



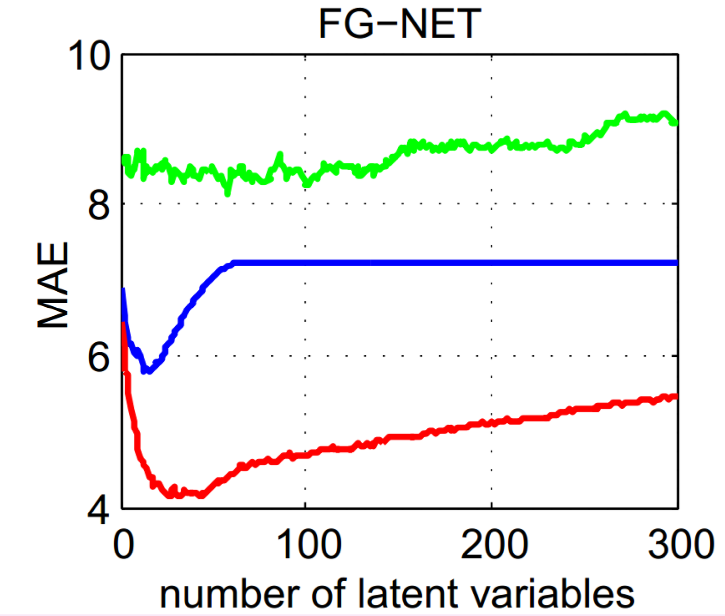
MORPH

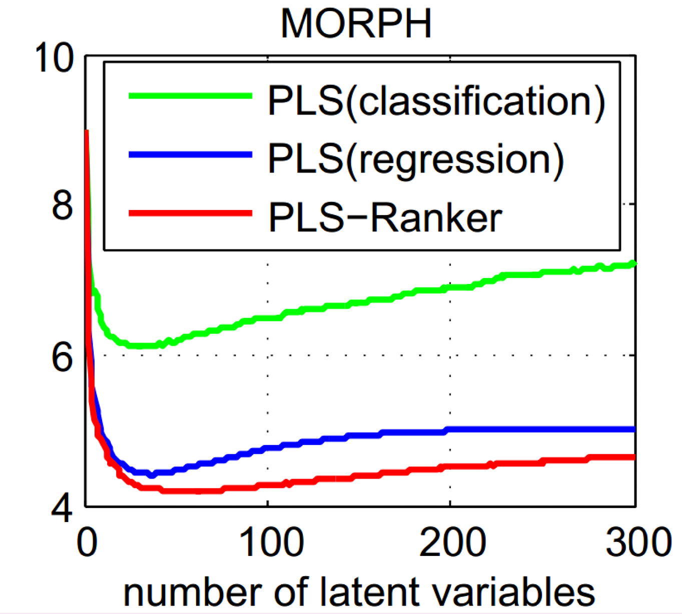


PLS-Ranker vs. PLS

- PLS can be used for classification and metric regression.
- Compare PLS-Ranker with the two usages of PLS on FG-NET and MORPH (Setting 1).

Method	MAE/year (num. of latent var.)	
	FGNET	MORPH
PLS (classification)	8.14 (58)	6.10 (33)
PLS (regression)	5.78 (17)	4.40 (37)
PLS-Ranker	**4.14** (45)	**4.17** (49)





Comparison with the state-of-the-art methods on FG-NET and MORPH (Setting 2)

5

1. MAE

Method	MAE/year	
	FGNET	MORPH
MTWGP [2]	4.83	6.28
PLO [14]	4.82	–
AAM+CA-SVR [15]	4.67	5.88
Feat. combine + select [22]	4.49	–
AAM+CS-OHRank [4]	4.48	6.07
Regularized CA-SVR [17]	4.37	–
Deep Feature+SVR [3]	4.26	4.77
AAM+PLS-Ranker	**4.14**	**5.38**
BIF+PLS-Ranker	–	**3.77**

2. Training time

Tested on an Intel(R) Core i5-3470 (3.2GHz), 8G RAM PC.

Method	Training time/min	
	FGNET	MORPH
OHRank [4]	1.30×10^4	3.02×10^4
SVR [23]	2.69×10^0	2.08×10^1
CA-SVR [15]	8.91×10^{-1}	6.10×10^0
AAM+PLS-Ranker	$\mathbf{7.20 \times 10^{-3}}$(0.43s)	$\mathbf{2.25 \times 10^{-2}}$(1.35s)
BIF+PLS-Ranker	–	$\mathbf{1.25 \times 10^{-1}}$(7.51s)

- Linear PLS is adopted.
- Simultaneous dimensionality reduction and ranking.
- Usually $p << N$ holds, and more latent variables (big p) are not necessarily in practice.

Robustness to race and gender variations

Results on the multi-source cross-race-and-gender age estimation problem

Train	Test	MAE/year (num. of latent var.)	
		CpDA [20]	PLS-Ranker
BF+WF	BM	6.47	**4.55** (27)
BF+WF	WM	5.70	**3.87** (49)
WM+BM	WF	6.58	**5.10** (80)
WM+BM	BF	6.40	**5.49** (67)
BF+BM	WF	6.59	**5.24** (88)
BF+BM	WM	5.23	**3.85** (63)
WF+WM	BF	6.32	**5.65** (39)
WF+WM	BM	5.96	**4.49** (31)
Average		5.99	**4.48 (25.31%)**

Please see our paper for **references** and other details **.**