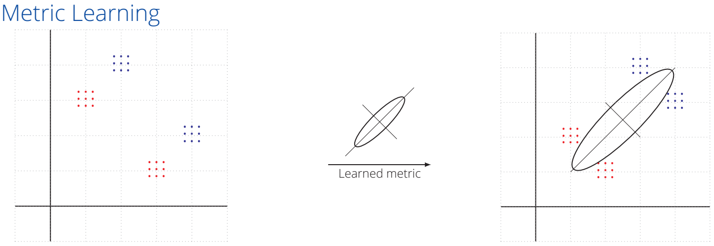
Scalable Metric Learning via Weighted Approximate Rank Component Analysis

Cijo Jose – François Fleuret
{cijo.jose,francois.fleuret}@idiap.ch

Overview

- WARCA is a new model for large-scale learning of Mahalanobis distances.
- WARCA optimizes the precision at top ranks by combining the WARP loss with a regularizer that favors orthonormal linear mappings and avoids rank-deficient embeddings.
- Non-linear WARCA through kernel-trick when data-set size permits kernel computation.
- Benchmarks on nine re-identification data-sets shows state-of-the-art performance both in terms of accuracy and speed.
- Experimental analysis also show how our new regularizer improves the performance.

Metric Learning



$$\mathcal{F}_W(x_i, x_j) := \|W(x_i - x_j)\|_2.$$

$$W^* = \underset{W}{\operatorname{argmin}} \frac{\lambda}{2}\Omega(W) + \mathbb{1}_{\mathcal{F}_W(\bullet,\bullet)-\mathcal{F}_W(\bullet,\bullet)\geq 0} \quad + \mathbb{1}_{\mathcal{F}_W(\bullet,\bullet)-\mathcal{F}_W(\bullet,\bullet)\geq 0}.$$

WARCA

Given data point / label pairs: $(x_n, y_n) \in \mathbb{R}^D \times \{1, \ldots, Q\}$, $n = 1, \ldots, N$. We define the following sets:

$$\mathcal{S} = \left\{(i,j) \in \{1, \ldots, N\}^2,\ y_i = y_j\right\}, \quad (1)$$

$$\mathcal{T}_y = \left\{k \in \{1, \ldots, N\},\ y_k \neq y\right\}. \quad (2)$$

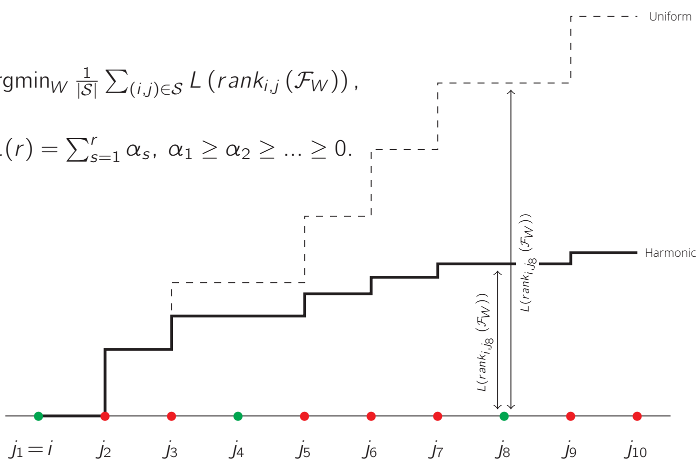
We define:

$$rank_{i,j}(\mathcal{F}_W) = \sum_{k \in \mathcal{T}_{y_i}} \mathbb{1}_{\mathcal{F}_W(x_i,x_k) \leq \mathcal{F}_W(x_i,x_j)}, \quad (3)$$

as the number of samples x_k of different labels which are closer to x_i than x_j. WARCA optimizes the WARP loss (Weston *et.al* IJCAI2011) on $rank_{i,j}(\mathcal{F}_W)$.

$$\operatorname{argmin}_W \frac{1}{|\mathcal{S}|}\sum_{(i,j)\in\mathcal{S}} L\left(rank_{i,j}(\mathcal{F}_W)\right),$$

$$L(r) = \sum_{s=1}^{r} \alpha_s,\ \alpha_1 \geq \alpha_2 \geq \ldots \geq 0.$$



Approximate OrthoNormal (AON) Regularizer

$$\|WW^T - \mathbf{I}\|^2. \quad (4)$$

- AON is simple and computationally efficient.
- AON achieves robustness and better generalization by preventing rank-deficient mappings.

Max-margin Reformulation

$$\underset{W}{\operatorname{argmin}} \frac{\lambda}{2}\|WW^T - \mathbf{I}\|^2 + \frac{1}{|\mathcal{S}|}\sum_{(i,j)\in\mathcal{S}}\sum_{k\in\mathcal{T}_{y_i}} L(rank^{\gamma}_{i,j}(\mathcal{F}_W))\frac{|\gamma + \xi_{ijk}|_+}{rank^{\gamma}_{i,j}(\mathcal{F}_W)}, \quad (5)$$

where:

$$\xi_{ijk} = \mathcal{F}_W(x_i, x_j) - \mathcal{F}_W(x_i, x_k) \quad (6)$$

and $rank^{\gamma}_{i,j}(\mathcal{F}_W)$ is the margin penalized rank:

$$rank^{\gamma}_{i,j}(\mathcal{F}_W) = \sum_{k\in\mathcal{T}_{y_i}} \mathbb{1}_{\gamma + \xi_{ijk} > 0}. \quad (7)$$

- Efficiently optimized using SGD.

Kernelization

$$\underset{A}{\operatorname{argmin}} \frac{\lambda}{2}\|AKA^T - \mathbf{I}\|^2 + \frac{1}{|\mathcal{S}|}\sum_{(i,j)\in\mathcal{S}}\sum_{k\in\mathcal{T}_{y_i}} L(rank^{\gamma}_{i,j}(\mathcal{F}_A))\frac{|\gamma + \xi_{ijk}|_+}{rank^{\gamma}_{i,j}(\mathcal{F}_A)}, \quad (8)$$

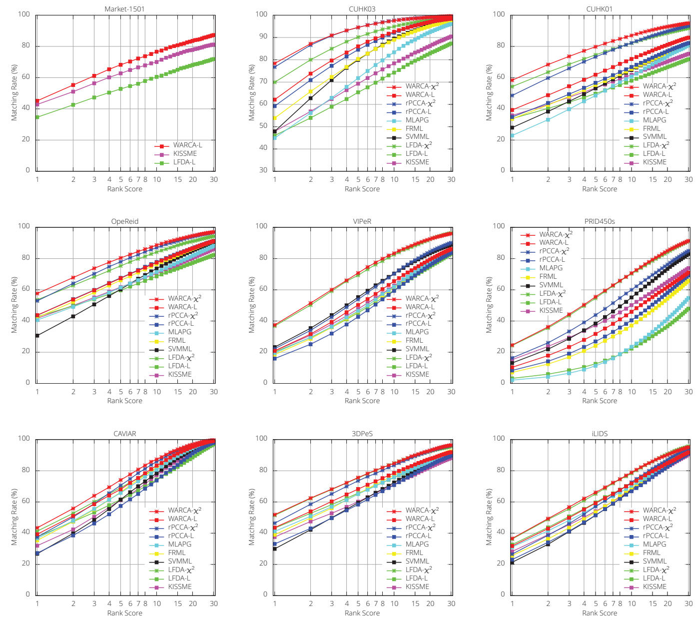
with:

$$\mathcal{F}_A(x_i, x_j) = \|AX^T(x_i - x_j)\|_2, \quad (9)$$

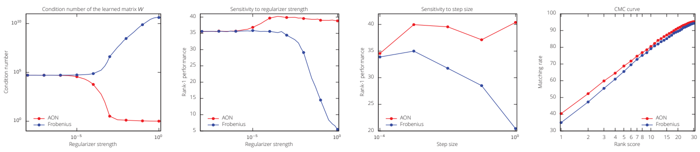
$$= \|A(\kappa_i - \kappa_j)\|_2. \quad (10)$$

Experimental Results

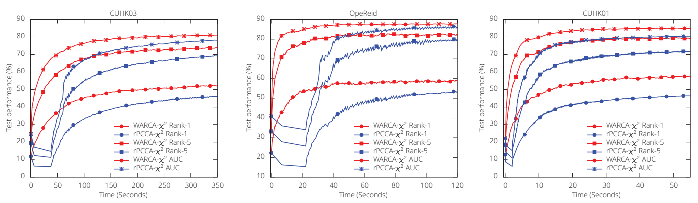
CMC Curves



Analysis of the AON regularizer



Analysis of the Training Time



Code available at: `https://github.com/idiap/warca`