Robust Face Recognition Using Online Learning

MICHAL VALKO, BRANISLAV KVETON, LING HUANG, DANIEL TING, MATTHAI PHILIPOSE

University of Pittsburgh

An example of a similarity graph over faces. The faces are vertices of the graph. The edges of the graph connect similar faces. Labeled faces are outlined by solid lines.

Goal
A face recognition algorithm that:
- Has a high accuracy
- Has a high recall
- Is robust to outliers
- Runs in real time

Approach
Online (real-time and incremental) learning of a similarity graph over observed faces and inference of face IDs based on the structure of the graph.

Algorithm (time step $t$):
- Remove outliers from the graph
- If the graph is too large, coarsen the adjacency graph by replacing the neighboring points with a set of local representative points
- Add the new face to the graph
- Infer the ID of the face based on the structure of the graph - a random walk that starts at the new face and terminates at the labeled faces

Challenges
High accuracy and a high recall are contradicting objectives. Achieving both with standard ML algorithms is typically impossible (unless the training set closely resembles the test set).

Adaptive ML algorithms can help to achieve this objective. The problem is that no explicit feedback (labels) is provided in real time.

Online Algorithm

Inputs:
an unlabeled example $x_t$
up to $n_g$ representative vertices $C_{t-1} = \{c_1, c_2, \ldots \}$
vertex multiplicities $v_{t-1}$

Algorithm:

$C_t = C_{t-1}$
$v_t = v_{t-1}$
while ($|C_t| = n_g + 1$)

$R = 2R$
greedily choose $C_t \subseteq C_{t-1}$ with $\min_{a \in C_t} \|a - b\| > R$
update $v_t$ based on the repartitioning if $x_t$ is closer than $R$ to any $c_i \in C_t$
$v_t(i) = v_t(i) + 1$
else
$v_t(|C_t| + 1) = 1$
add $x_t$ to the position ($|C_t| + 1$) in $C_t$
build a similarity matrix $W_t$ over the vertices $C_t$
build a matrix $V_t$ whose diagonal elements are $v_t$
compute the Laplacian $\hat{L}$ of the graph $W_t$
infer labels on the graph:

$\hat{y}_t = \arg \min_{\hat{y}} \langle \hat{L} + \gamma V_t \rangle \hat{y}$
s.t. $L_t = y_t$ for all labeled examples up to the time $t$ make a prediction $\hat{y}_t = \text{sgn}(\hat{L}_t \hat{y}_t)$

Outputs:
a prediction $\hat{y}_t$
up to $n_g$ representative vertices $C_t = \{c_1, c_2, \ldots \}$
vertex multiplicities $v_t$

Online harmonic solution at the time step $t$. The main parameters of the algorithm is the regularizer $\gamma_t$ and the maximum number of vertices $n_g$.

Prediction Error Analysis

$\frac{1}{n} \sum_{t=1}^{n} (\hat{y}_t - y_t)^2$ vs $\frac{9}{2n} \sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2$
$+$ $\frac{9}{2n} \sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2$
$+$ $\frac{9}{2n} \sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2$

True risk close to empirical by the algorithm stability argument of Cortes et al. (2008)
Difference between the offline and online prediction
Quality of quantization

Similarity Matrix
- Defined over set of faces, higher weights to the pixels in the center
- $w_{ij} = \exp \left[ -\frac{d(x_i, x_j)}{2 \sigma^2} \right]$
  where $d(x_i, x_j) = \min \left\{ \|x_i - x_j\|_2, \|x_i - x_j\|_2, \|x_i - x_j\|_2 \right\}$

Data Quantization
- Cannot store all the past data
- Similarity graph needs to be reasonably small
- Use $k$-centers algorithm to maintain constant graph size
- Represent the multiple nodes by a single one
- Keep track of multiplicities $\hat{e}_t = (\hat{L} + \gamma V)^{-1} W_t \hat{e}_t$

Regularized Harmonic Solution
Minimum satisfies the harmonic property and has a closed form solution.

Robustness to outliers
Nearest Neighbor
Nearest Neighbor
Our Algorithm
Online Semi-Supervised Boosting
Snapshots from the Adaptation dataset. 3 locations and 2 camera positions.

Nearest Neighbor
Nearest Neighbor