Face Recognition Using Random Walks on Graphs: Real-Time Learning without Explicit Feedback

Michal Valko (Computer Science Department)
Branislav Kveton, Ling Huang, Daniel Ting
Main Ideas

• **Goal:** Adaptation to patterns with minimal human feedback (labels)
  – Most of data around is unlabeled
  – Labeling is expensive

• **Solution:** Semi – Supervised learning (Machine Learning)
  – Labeled examples are provided in the beginning
    • Provide initial bias
  – Unlabeled examples come as available

• **Approach:** Regularized graph–based inference + quantization
Semi-supervised learning
Semi-supervised learning

Data

Supervised

Semi-Supervised
Face Similarities
Face Similarities
Face Similarities

\[
d(x_i, x_j) = \min \left\{ \| x_i - x_j \|_{2,\psi}, \| (x_i - \bar{x}_i) - (x_j - \bar{x}_j) \|_{2,\psi}, \| x_i / \bar{x}_i - x_j / \bar{x}_j \|_{2,\psi} \right\}
\]
Graph from faces
Graph-based
Semi-Supervised Learning

\[ \ell^* = \arg \min_{\ell} \ell^T L \ell \]
Harmonic Function Solution (HFS)

- Labels of unlabeled vertices are inferred using the harmonic function solution
Dealing with Outliers
Regularized HFS

\[ \ell^* = \arg \min_{\ell} \ell^T (\mathbf{g}^T I + L) \ell \]

s.t. \( \ell_i = y_i \) for all \( i \in \ell \);
Regularization

Low confidence

$\gamma_g = 100.000$

$\gamma_g = 1.000$

$\gamma_g = 0.010$

ONLY LABELED

ALL DATA
Online HFS

**Inputs:** an example $x_t$, a data adjacency graph $W$

**Algorithm:**
Add $x_t$ to the graph $W$ and compute the Laplacian $L$

Infer labels on the graph:

$$
\min_{\lambda \in \mathbb{R}^N} \lambda^T (L + \gamma I) \lambda \quad \text{s.t. } \lambda_i = y_i \text{ for all } i \in l
$$

Predict $\hat{y}_t = \lambda_t$

**Outputs:** a prediction $\hat{y}_t$, an updated data adjacency graph $W$

What is wrong with this algorithm?
Online HFS

**Inputs:** an example $x_t$, a data adjacency graph $W$

**Algorithm:**

- If the graph $W$ has more than $M$ vertices, quantize it
- Add $x_t$ to the graph $W$ and compute the Laplacian $L$
- Infer labels on the graph:
  $$\min_{\lambda \in \mathbb{R}^N} \lambda^T (L + \gamma I)\hat{\lambda} \quad \text{s.t.} \quad \lambda_i = y_i \text{ for all } i \in l$$
  $$\text{Predict } \hat{y}_t = \lambda_t$$

**Outputs:** a prediction $\hat{y}_t$, an updated data adjacency graph $W$
Incremental k-centers
Incremental k-centers
Incremental k-centers
Demostration
Theoretical Guarantees

• We seek a regret bound of the form:

\[
\frac{1}{N} \sum_t (\hat{y}_t - y_t)^2 \leq \frac{1}{N} \sum_t (y^*_t - y_t)^2 + \frac{1}{N} \sum_t (y'_t - y^*_t)^2 + \frac{1}{N} \sum_t (\hat{y}_t - y'_t)^2
\]

• The errors should be bounded on the order of \(O(\sqrt{N})\)
OfficeSpace Dataset

Snapshots

• 8 people
• Only 4 faces are labeled
Results (OfficeSpace)

Our method

Nearest Neighbor

Better

Precision [%]

Recall [%]
Adaptation Dataset

• 3 locations, different light conditions
• 8 camera positions
Results (Adaptation)

- Our method
- Nearest Neighbor
- Online Semi-Supervised Boosting
Conclusions

• Algorithm for semi-supervised learning
  – Takes advantage of the manifold structure in the data
• Requires minimal feedback
  – Only 1 or few labeled examples
• Works online and requires constant storage
• Theoretical guarantees on success rates of our methods
• Future work:
  – other data reduction methods
  – other domains: object recognition, augmented reality