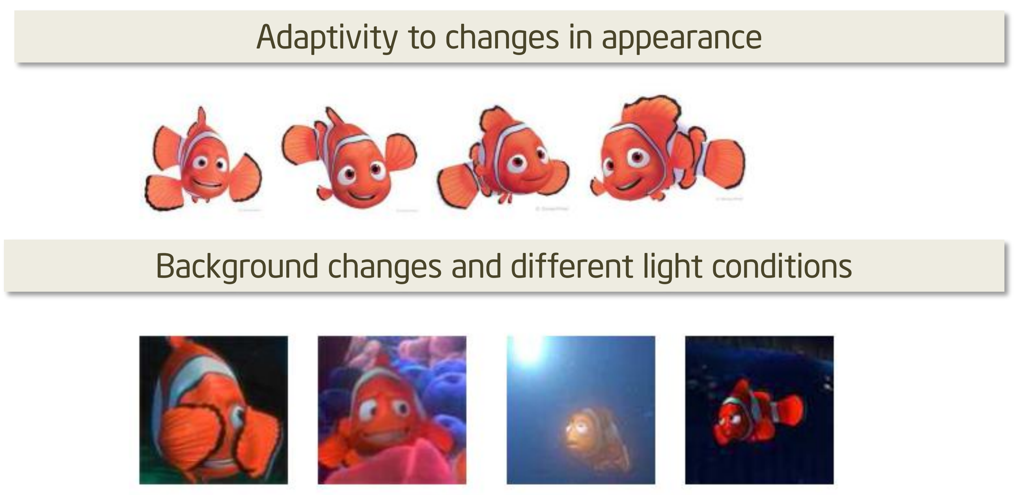
Online Semi-Supervised Learning on Quantized Graphs

MICHAL VALKO, BRANISLAV KVETON, LING HUANG, DANIEL TING



Grabner et. al (2008)

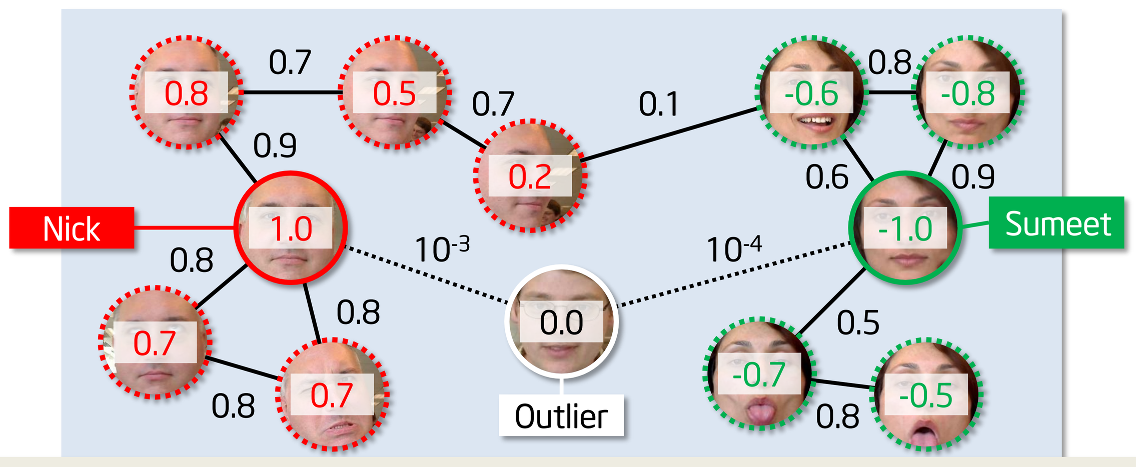
Goal

Vision applications need algorithms that:

- Require minimal human feedback
- Adapt to changes in environment
- Are robust to outliers
- Run in real time
- Have high accuracy
- Have high recall

Approach

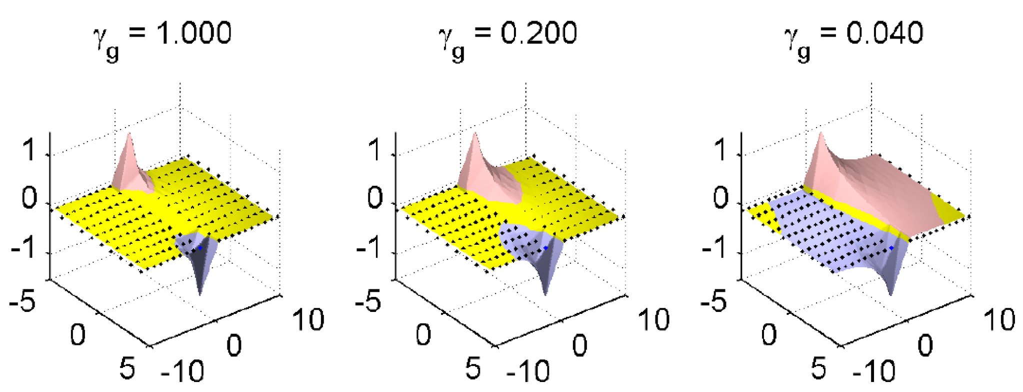
These requirements fit the paradigm of **Online Semi-Supervised Learning**. We perform online (real-time and incremental) learning of a similarity graph over observed examples and do inference based on the structure of the graph.



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.

Background

- Labels of unlabeled vertices are inferred using the harmonic solution.
- Regularization controls the amount of extrapolation to unlabeled data. The smaller the regularizer, the more we trust unlabeled data.



Online Learning

- Cannot store all the past data
- Similarity graph needs to be reasonably small due to computational complexity.

Data Quantization

- Use *k*-centers algorithm by Charikar et. al (1997) to maintain constant graph size.
- Represent the multiple nodes by a single centroid.
- Keep track of *multiplicities.*

Online Algorithm for Quantized Harmonic Solution

Online harmonic solution at the time step *t*. The main parameters of the algorithm is the regularizer γ_g and the maximum number of centroids *k*.

Inputs:

example x_t
data adjacency graph W^q of at most *k* vertices

Algorithm:

Add the example x_t to the graph W^q and quantize it
Compute the Laplacian L^q of W^q and infer labels:

$$\ell^q[t] \leftarrow \operatorname{argmin}_\ell \ell^T (L^q + \gamma_g V_t)\ell$$
$$\text{s.t. } \ell_i = y_i \text{ for all } i \in l$$

Predict $\hat{y}_t = \operatorname{sgn}(\ell_t^q[t])$

Outputs:

a prediction $\hat{y}_t$
an updated data adjacency graph W^q

Prediction Error Analysis

Our learner solves an online regression problem. The error we want to minimize is bounded as:

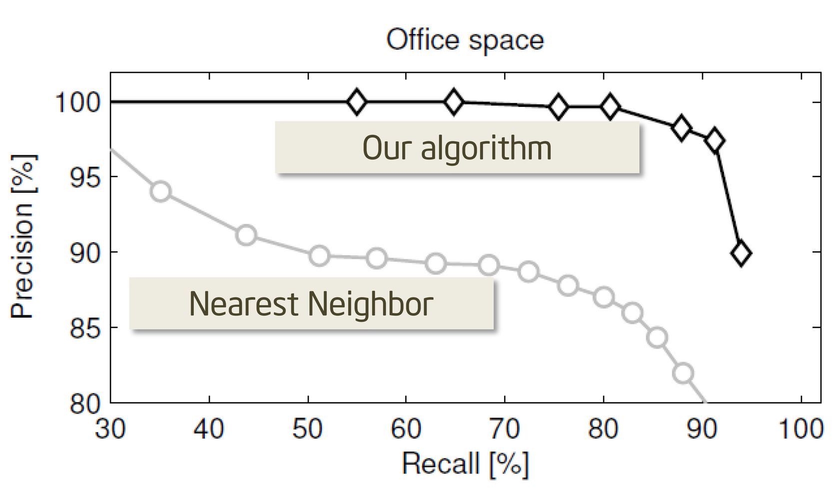
$$\frac{1}{n}\sum_{t=1}^{n}(\ell_t^{oq}[t]-y_t)^2 \leq \frac{9}{2n}\sum_{t=1}^{n}(\ell_t^*-y_t)^2 + \frac{9}{2n}\sum_{t=1}^{n}(\ell_t^o[t]-\ell_t^*)^2 + \frac{9}{2n}\sum_{t=1}^{n}(\ell_t^{oq}[t]-\ell_t^o[t])^2$$

- **True risk close to empirical risk.** Regularized harmonic solution is a special case of the unconstrained regularization. By the algorithm stability argument of Cortes et al. (2008) we can bound the its generalization error.
- **Difference between the offline and online prediction.** When the Laplacians of the both online and offline similarity graph are regularized enough then the corresponding solutions are close to zero and therefore close to each other.
- **Quality of quantization.** This error bounds how much can harmonic solution change when we approximate the graph of all previously seen examples with its quantized version. Our bound on the difference is linear in the Frobenius norm of the difference of the corresponding Laplacians $\|L^q - L^o\|_F$ which is o(1) under certain conditions.
- Combining the three terms above we can prove the following regret bound:

$$\frac{1}{n}\sum_{t=1}^{n}(\ell_t^q[t]-y_t)^2 \leq \frac{9}{2n_l}\sum_{i\in l}(\ell_i^*-y_i)^2 + O(n^{-\frac{1}{2}})$$

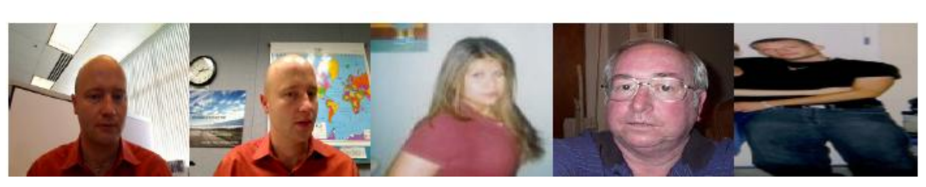
Office space dataset

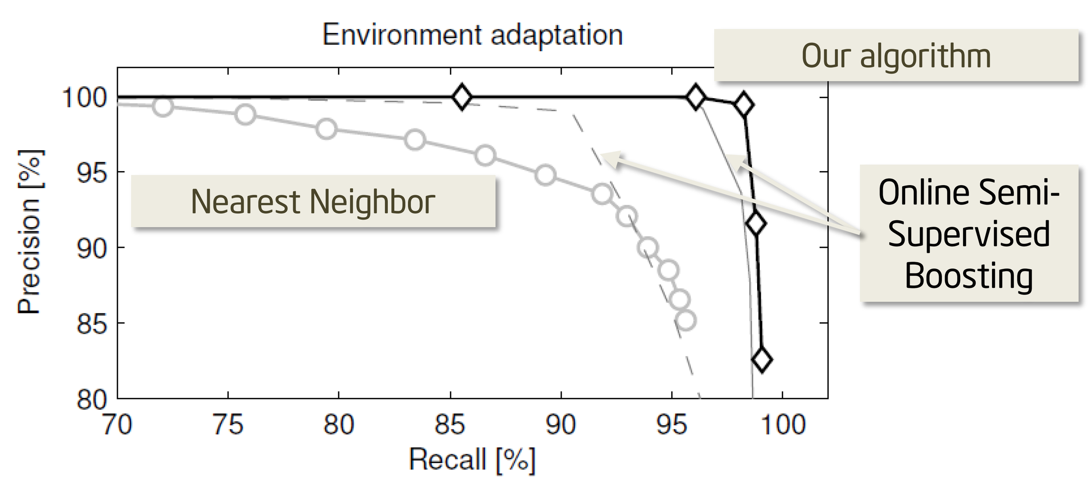
- Multi-class: 8 people who walk in front of a camera and make funny faces.
- When a person shows up on the camera for the first time, we label four faces of the person.



Environment adaptation dataset

- Faces of a single person, which are captured at various locations.
- The first four faces in the cubicle are labeled.
- To test the sensitivity of the recognizer to outliers, we appended the dataset by random faces.





UCI Letter recognition

- Left: We fix the number of centroids at 200. As time t increases, the error $\|L^q-L^o\|_F$ slowly levels off.
- Right: We fix the learning time at t = n and vary the number of centroids.