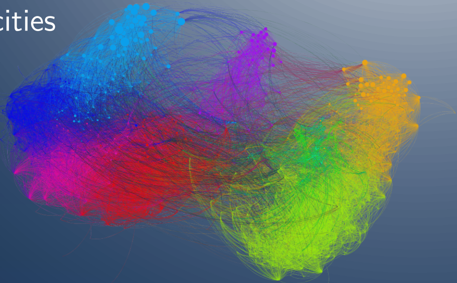
Graphs in Machine Learning

Online SSL: Graph Quantization

Charikar's Algorithm and Multiplicities



Michal Valko

Inria & ENS Paris-Saclay, MVA

Partially based on material by: Branislav Kveton, Mikhail Belkin, Jerry Zhu

Online SSL with Graphs: Graph Quantization

An idea: incremental k-centers

Online SSL with Graphs: Graph Quantization

An idea: incremental k-centers

Doubling algorithm of Charikar et al. Charikar et al., 1997

Online SSL with Graphs: Graph Quantization

An idea: incremental *k*-centers

Doubling algorithm of Charikar et al. Charikar et al., 1997

Keeps up to *k* centers $C_t = \{\mathbf{c}_1, \mathbf{c}_2, \dots\}$ with

Online SSL with Graphs: Graph Quantization

An idea: incremental *k*-centers

Doubling algorithm of Charikar et al. Charikar et al., 1997

Keeps up to k centers $C_t = \{\mathbf{c}_1, \mathbf{c}_2, \dots\}$ with

- Distance $\mathbf{c}_i, \mathbf{c}_j \in C_t$ is at least $\geq R$

Online SSL with Graphs: Graph Quantization

An idea: incremental k-centers

Doubling algorithm of Charikar et al. Charikar et al., 1997

Keeps up to k centers $C_t = \{\mathbf{c}_1, \mathbf{c}_2, \dots\}$ with

- Distance $\mathbf{c}_i, \mathbf{c}_j \in C_t$ is at least $\geq R$
- For each new $\mathbf{x}_t$, distance to some $\mathbf{c}_i \in C_t$ is less than R.

Online SSL with Graphs: Graph Quantization

An idea: incremental k-centers

Doubling algorithm of Charikar et al. Charikar et al., 1997

Keeps up to k centers $C_t = \{\mathbf{c}_1, \mathbf{c}_2, \dots\}$ with

- Distance $\mathbf{c}_i, \mathbf{c}_j \in C_t$ is at least $\geq R$
- For each new $\mathbf{x}_t$, distance to some $\mathbf{c}_i \in C_t$ is less than R.
- $|C_t| \leq k$

Online SSL with Graphs: Graph Quantization

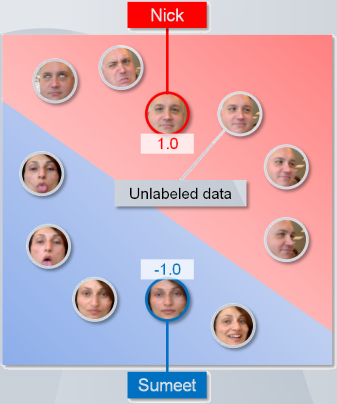
An idea: incremental k-centers

Doubling algorithm of Charikar et al. Charikar et al., 1997

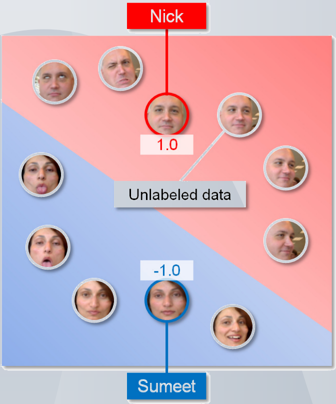
Keeps up to k centers $C_t = \{\mathbf{c}_1, \mathbf{c}_2, \dots\}$ with

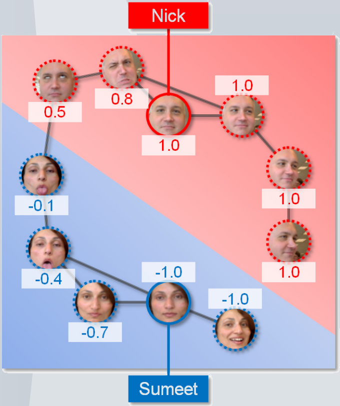
- Distance $\mathbf{c}_i, \mathbf{c}_j \in C_t$ is at least $\geq R$
- For each new $\mathbf{x}_t$, distance to some $\mathbf{c}_i \in C_t$ is less than R.
- $|C_t| \leq k$
- if not possible, R is doubled

Online SSL with Graphs: Graph Quantization

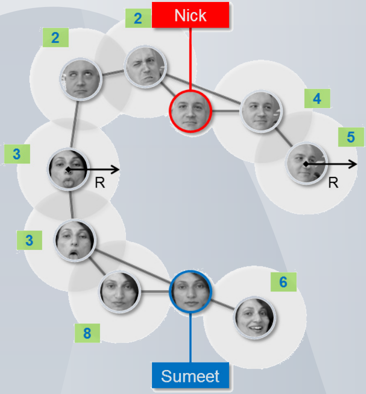


Online SSL with Graphs: Graph Quantization

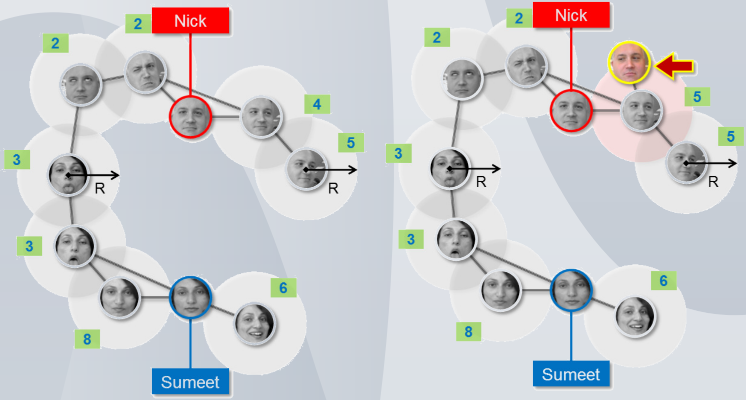




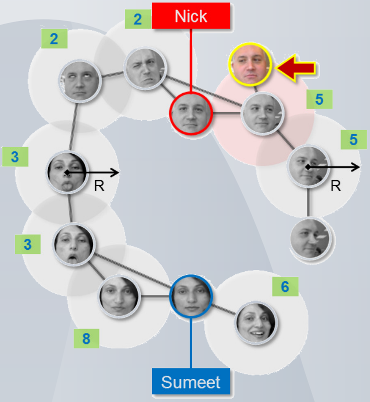
Online SSL with Graphs: Graph Quantization



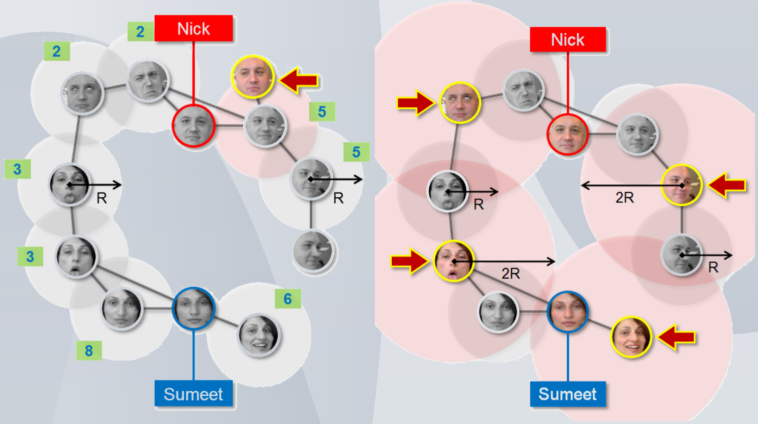
Online SSL with Graphs: Graph Quantization



Online SSL with Graphs: Graph Quantization



Online SSL with Graphs: Graph Quantization



Online SSL with Graphs: Graph Quantization

Online k-centers

1: an unlabeled $\mathbf{x}_t$, a set of centroids C_{t-1}, multiplicities $\mathbf{v}_{t-1}$

Online SSL with Graphs: Graph Quantization

Online k-centers

1: an unlabeled $\mathbf{x}_t$, a set of centroids C_{t-1}, multiplicities $\mathbf{v}_{t-1}$
2: **if** $(|C_{t-1}| = k + 1)$ **then**
3: $\quad R \leftarrow mR$

Online SSL with Graphs: Graph Quantization

Online k-centers

```
an unlabeled x_t, a set of centroids C_{t-1}, multiplicities v_{t-1}
if (|C_{t-1}| = k + 1) then
    R ← mR
    greedily repartition C_{t-1} into C_t such that:
        no two vertices in C_t are closer than R
        for any c_i ∈ C_{t-1} exists c_j ∈ C_t such that d(c_i, c_j) < R
```

Online SSL with Graphs: Graph Quantization

Online k-centers

an unlabeled $\mathbf{x}_t$, a set of centroids C_{t-1}, multiplicities $\mathbf{v}_{t-1}$
if $(|C_{t-1}| = k+1)$ **then**
  $R \leftarrow mR$
  greedily repartition C_{t-1} into C_t such that:
    no two vertices in C_t are closer than R
    for any $\mathbf{c}_i \in C_{t-1}$ exists $\mathbf{c}_j \in C_t$ such that $d(\mathbf{c}_i, \mathbf{c}_j) < R$
  update $\mathbf{v}_t$ to reflect the new partitioning
else
  $C_t \leftarrow C_{t-1}$
  $\mathbf{v}_t \leftarrow \mathbf{v}_{t-1}$

Online SSL with Graphs: Graph Quantization

Online k-centers

an unlabeled $\mathbf{x}_t$, a set of centroids C_{t-1}, multiplicities $\mathbf{v}_{t-1}$
if $(|C_{t-1}| = k + 1)$ **then**
    $R \leftarrow mR$
    greedily repartition C_{t-1} into C_t such that:
        no two vertices in C_t are closer than R
        for any $\mathbf{c}_i \in C_{t-1}$ exists $\mathbf{c}_j \in C_t$ such that $d(\mathbf{c}_i, \mathbf{c}_j) < R$
    update $\mathbf{v}_t$ to reflect the new partitioning
else
    $C_t \leftarrow C_{t-1}$
    $\mathbf{v}_t \leftarrow \mathbf{v}_{t-1}$
end if
if $\mathbf{x}_t$ is closer than R to any $\mathbf{c}_i \in C_t$ **then**
    $\mathbf{v}_t(i) \leftarrow \mathbf{v}_t(i) + 1$
else

Online SSL with Graphs: Graph Quantization

Online k-centers

an unlabeled $\mathbf{x}_t$, a set of centroids C_{t-1}, multiplicities $\mathbf{v}_{t-1}$
if $(|C_{t-1}| = k + 1)$ **then**
  $R \leftarrow mR$
  greedily repartition C_{t-1} into C_t such that:
    no two vertices in C_t are closer than R
    for any $\mathbf{c}_i \in C_{t-1}$ exists $\mathbf{c}_j \in C_t$ such that $d(\mathbf{c}_i, \mathbf{c}_j) < R$
  update $\mathbf{v}_t$ to reflect the new partitioning
else
  $C_t \leftarrow C_{t-1}$
  $\mathbf{v}_t \leftarrow \mathbf{v}_{t-1}$
end if
if $\mathbf{x}_t$ is closer than R to any $\mathbf{c}_i \in C_t$ **then**
  $\mathbf{v}_t(i) \leftarrow \mathbf{v}_t(i) + 1$
else
  $\mathbf{v}_t(|C_t| + 1) \leftarrow 1$
end if

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

C_t is changing. How far can $\mathbf{x}$ be from some $\mathbf{c}$?

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

C_t is changing. How far can $\mathbf{x}$ be from some $\mathbf{c}$?

R

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

C_t is changing. How far can $\mathbf{x}$ be from some $\mathbf{c}$?

$$R + \frac{R}{R}$$

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

C_t is changing. How far can $\mathbf{x}$ be from some $\mathbf{c}$?

$$R + \frac{R}{R} + \frac{R}{R^2}$$

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

C_t is changing. How far can $\mathbf{x}$ be from some $\mathbf{c}$?

$$R + \frac{R}{R} + \frac{R}{R^2} + \cdots = R\left(1 + \frac{1}{R} + \frac{1}{R^2} + \cdots\right)$$

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

C_t is changing. How far can $\mathbf{x}$ be from some $\mathbf{c}$?

$$R + \frac{R}{R} + \frac{R}{R^2} + \cdots = R\left(1 + \frac{1}{R} + \frac{1}{R^2} + \cdots\right) = \frac{RR}{R-1}$$

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

C_t is changing. How far can $\mathbf{x}$ be from some $\mathbf{c}$?

$$R + \frac{R}{R} + \frac{R}{R^2} + \cdots = R\left(1 + \frac{1}{R} + \frac{1}{R^2} + \cdots\right) = \frac{RR}{R-1}$$

Guarantees

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

C_t is changing. How far can $\mathbf{x}$ be from some $\mathbf{c}$?

$$R + \frac{R}{R} + \frac{R}{R^2} + \cdots = R\left(1 + \frac{1}{R} + \frac{1}{R^2} + \cdots\right) = \frac{RR}{R-1}$$

Guarantees: 8-approximation algorithm.

Online SSL with Graphs: Graph Quantization

Doubling algorithm Charikar et al., 1997

To reduce growth of R, we use $R \leftarrow R \times R$, with $R \geq 1$

C_t is changing. How far can $\mathbf{x}$ be from some $\mathbf{c}$?

$$R + \frac{R}{R} + \frac{R}{R^2} + \cdots = R\left(1 + \frac{1}{R} + \frac{1}{R^2} + \cdots\right) = \frac{RR}{R-1}$$

Guarantees: 8-approximation algorithm.

Why not incremental k-means?

Online SSL with Graphs

Video examples

http://www.bkveton.com/videos/Coffee.mp4

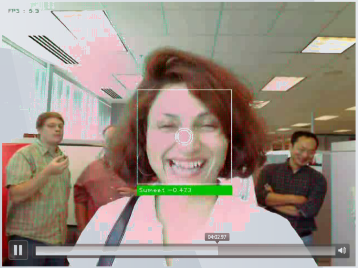
http://www.bkveton.com/videos/Ad.mp4

https://misovalko.github.io/publications/kveton2009nipsdemo.adaptation.mov

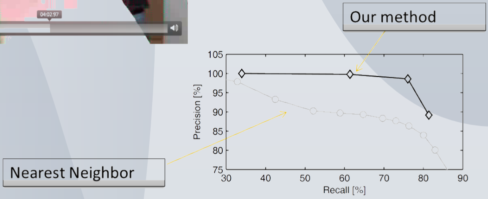
https://misovalko.github.io/publications/kveton2009nipsdemo.officespace.mov

https://misovalko.github.io/publications/press-intel-2015.mp4

SSL with Graphs: Some experimental results



- 8 people classification
- Making funny faces
- 4 faces/person are labeled



Michal Valko

michal.valko@inria.fr

Inria & ENS Paris-Saclay, MVA

https://misovalko.github.io/mva-ml-graphs.html