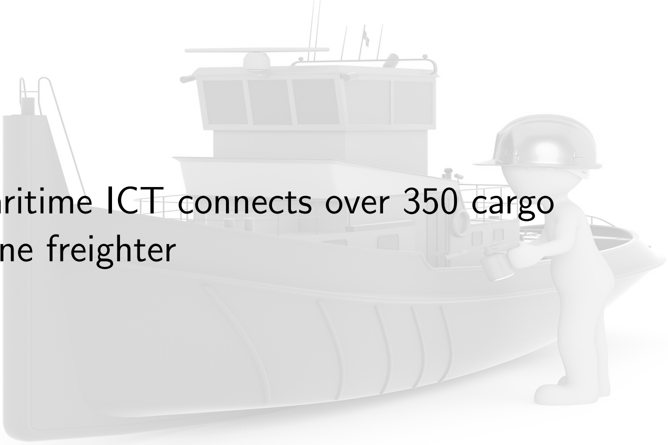
Summary Extraction on Data Streams in Embedded Systems

Sebastian Buschjäger and Katharina Morik

TU Dortmund University - Computer Science - Artificial Intelligence Group

September 18, 2017

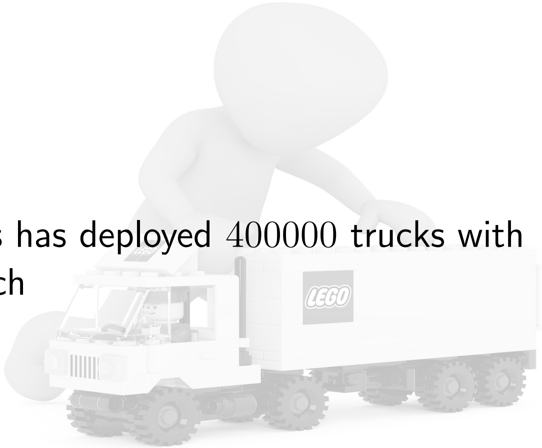
So... IoT hype?!



2016

Ericsson Maritime ICT connects over 350 cargo vessels on one freighter

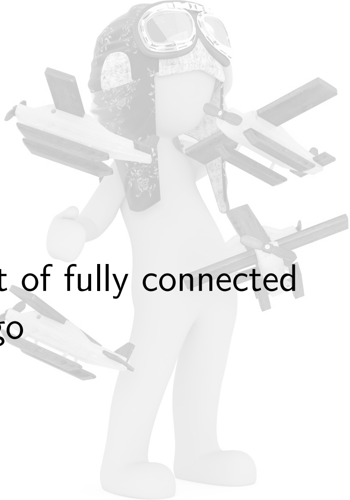
So... IoT hype?!



2016

Daimler Trucks has deployed 400000 trucks with 400 sensors each

So... IoT hype?!



2016

Virgin Atlantic announces fleet of fully connected Boeing 787 machines and cargo

IoT means large autonomous systems

Common intuition

There will be more devices

We will get more data

Systems will become more autonomous

IoT means large autonomous systems

Common intuition

There will be more devices

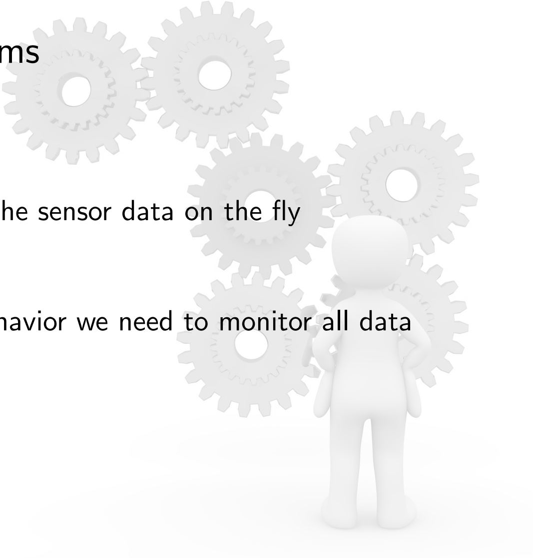
We will get more data

Systems will become more autonomous

Question

What to do if something unexpected happens?

Goal Monitor systems



Clear

Nobody can monitor all the sensor data on the fly

But

To detect unexpected behavior we need to monitor all data

Goal Monitor systems

Clear

Nobody can monitor all the sensor data on the fly

But

To detect unexpected behavior we need to monitor all data

Idea

Compute summaries on the fly while sensor data is generated

Goal Monitor systems

Then

Human expert can inspect summaries
Perform operations on summary etc.

Goal Monitor systems

Then

Human expert can inspect summaries
Perform operations on summary etc.

Constraint

Different data types + theoretically sound

Data summarization Some theory

Intuition

Use set function f to measures expressiveness of summary S

Goal

$$\max_{S \subseteq V, |S| \leq k} f(S)$$

Data summarization Some theory

Intuition
Use set function f to measures expressiveness of summary S

Goal

$$\max_{S\subseteq V, |S|\leq k} f(S)$$

Gain Let $f: V \to \mathbb{R}$ and let $e \in V$ and $S \subseteq V$:

$$\Delta_f(e|S) = f(S \cup \{e\}) - f(S)$$

Summarization Sieve-Streaming

Badanidiyuru et al. 2014 Sieve-Streaming
Item e arrives one at a time
Immediately decide if e should be added to summary

Summarization Sieve-Streaming

Badanidiyuru et al. 2014 Sieve-Streaming
Item e arrives one at a time
Immediately decide if e should be added to summary

Idea Introduce novelty threshold v. Add e if

$$\Delta_f(e|S) > v$$

Summarization Sieve-Streaming

Badanidiyuru et al. 2014 Sieve-Streaming
Item e arrives one at a time
Immediately decide if e should be added to summary

Idea Introduce novelty threshold v. Add e if

$$\Delta_f(e|S) > v$$

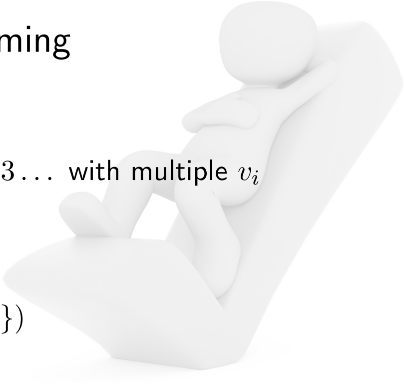
Challenge What is the “optimal” v?

Summarization Sieve-Streaming

Idea

Manage multiple summaries $i = 1, 2, 3\ldots$ with multiple v_i
$\rightarrow$ "sieve" out unimportant elements

Summarization Sieve-Streaming



Idea

Manage multiple summaries $i = 1, 2, 3 \ldots$ with multiple v_i
$\rightarrow$ "sieve" out unimportant elements

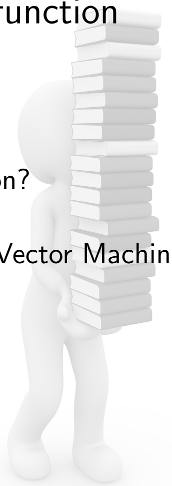
By sumodularity

$v_i \in [m, km]$ with $m = \max_{e \in V} f(\{e\})$
Then solution is $\frac{1}{2} - \varepsilon$ approximation

Note

This is independent from f

Submodular maximization The right function



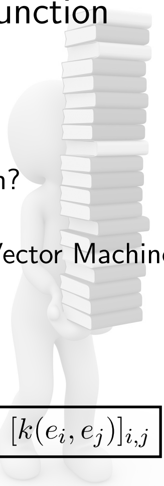
Question

What submodular function f captures summarization?

Herbrich et al. 2003 / Seeger 2004 Informative Vector Machine

$$f(S) = \frac{1}{2} \log \det \left(\mathcal{I} + \sigma^{-2} \Sigma_S \right)$$

Submodular maximization The right function



Question

What submodular function f captures summarization?

Herbrich et al. 2003 / Seeger 2004 Informative Vector Machine

$$f(S) = \frac{1}{2} \log \det \left(\mathcal{I} + \sigma^{-2} \Sigma_S \right)$$

$k \times k$ identiy matrix

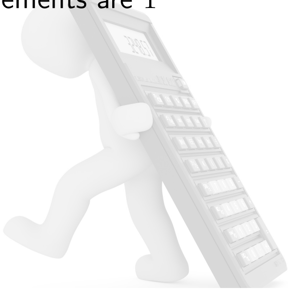
kernel matrix $K = [k(e_i, e_j)]_{i,j}$

IVM for data summarization

Since we know f, reduce interval!

Note Assume $k(e_i, e_i) = 1$

Least-expressive summary All off-diagonal elements are 1



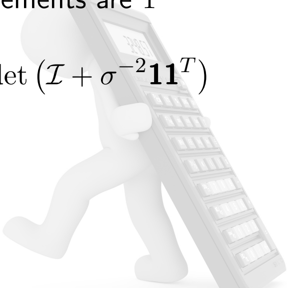
IVM for data summarization

Since we know f, reduce interval!

Note Assume $k(e_i, e_i) = 1$

Least-expressive summary All off-diagonal elements are 1

$$f(S) = \frac{1}{2} \log\det\left(\mathcal{I} + \sigma^{-2}\Sigma_S\right) = \frac{1}{2} \log\det\left(\mathcal{I} + \sigma^{-2}\mathbf{1}\mathbf{1}^T\right)$$

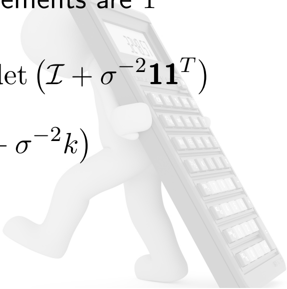


IVM for data summarization

Since we know f, reduce interval!

Note Assume $k(e_i, e_i) = 1$

Least-expressive summary All off-diagonal elements are 1



$$
\begin{aligned}
f(S) &= \frac{1}{2}\log\det\left(\mathcal{I} + \sigma^{-2}\Sigma_S\right) = \frac{1}{2}\log\det\left(\mathcal{I} + \sigma^{-2}\mathbf{1}\mathbf{1}^T\right) \\
&= \frac{1}{2}\log\left(1 + \sigma^{-2}\mathbf{1}^T\mathbf{1}\right) = \frac{1}{2}\log\left(1 + \sigma^{-2}k\right)
\end{aligned}
$$

IVM for data summarization

Since we know f, reduce interval!

Note Assume $k(e_i, e_i) = 1$

Most-expressive summary All off-diagonal elements are 0

IVM for data summarization

Since we know f, reduce interval!

Note Assume $k(e_i, e_i) = 1$

Most-expressive summary All off-diagonal elements are 0

$$f(S) = \frac{1}{2} \log \det \left(\mathcal{I} + \sigma^{-2} \Sigma_S \right) = \frac{1}{2} \log \det \left(\mathcal{I}(1 + \sigma^{-2}) \right)$$

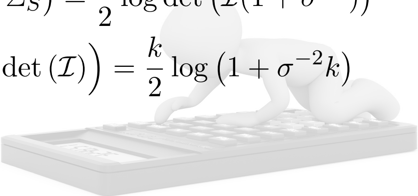
IVM for data summarization

Since we know f, reduce interval!

Note Assume $k(e_i, e_i) = 1$

Most-expressive summary All off-diagonal elements are 0

$$
\begin{aligned}
f(S) &= \frac{1}{2}\log\det\left(\mathcal{I} + \sigma^{-2}\Sigma_S\right) = \frac{1}{2}\log\det\left(\mathcal{I}(1+\sigma^{-2})\right) \\
&= \frac{1}{2}\log\left((1+\sigma^{-2})^k \det\left(\mathcal{I}\right)\right) = \frac{k}{2}\log\left(1+\sigma^{-2}k\right)
\end{aligned}
$$



Sieve-Streaming enhancements

Result

Number of sieves reduced without performance loss

Sieve-Streaming enhancements

Result

Number of sieves reduced without performance loss

Default $v_i \in [\frac{1}{2}\log(1+\sigma^{-2}), \frac{k}{2}\log(1+\sigma^{-2})]$

Reduced $v_i \in [\frac{1}{2}\log(1+k\sigma^{-2}), \frac{k}{2}\log(1+\sigma^{-2})]$

Sieve-Streaming enhancements

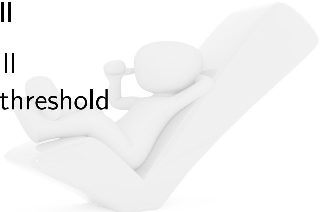
Result

Number of sieves reduced without performance loss

Default $v_i \in [\frac{1}{2}\log(1+\sigma^{-2}), \frac{k}{2}\log(1+\sigma^{-2})]$
Reduced $v_i \in [\frac{1}{2}\log(1+k\sigma^{-2}), \frac{k}{2}\log(1+\sigma^{-2})]$

More improvements Reopen sieves once full

Sieves with small threshold will quickly be full
Save summary, and reopen sieve with larger threshold



Sieve-Streaming enhancements

Result

Number of sieves reduced without performance loss

Default $v_i \in [\frac{1}{2}\log(1+\sigma^{-2}), \frac{k}{2}\log(1+\sigma^{-2})]$
Reduced $v_i \in [\frac{1}{2}\log(1+k\sigma^{-2}), \frac{k}{2}\log(1+\sigma^{-2})]$

More improvements Reopen sieves once full

Sieves with small threshold will quickly be full
Save summary, and reopen sieve with larger threshold

$\Rightarrow$ Increase utility value with same number of sieves

Experiments Questions

Question 1 Are summaries with IVM really expressive?

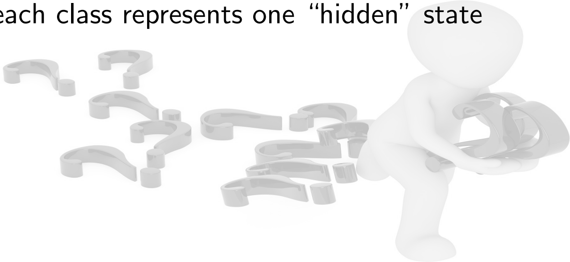
Experiments Questions

Question 1 Are summaries with IVM really expressive?

→ Summaries should contain “hidden” states of data

→ Extract summary of classification task

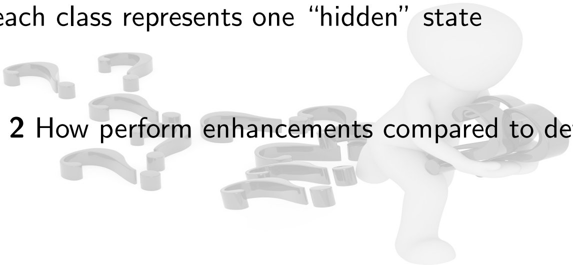
→ Then each class represents one “hidden” state



Experiments Questions

Question 1 Are summaries with IVM really expressive?

→ Summaries should contain "hidden" states of data
→ Extract summary of classification task
→ Then each class represents one "hidden" state



Question 2 How perform enhancements compared to default?

Experiments Data

Synthetic data

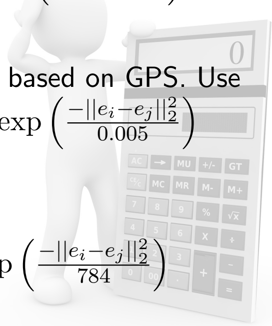
GMM with 4 dimensions and 4 classes. Use
$K = 10, \ldots, 24,\ \varepsilon = 0.1,\ \sigma = 1,\ k(e_i, e_j) = \exp\left(\frac{-||e_i - e_j||_2^2}{10}\right)$

UJIndoor Location

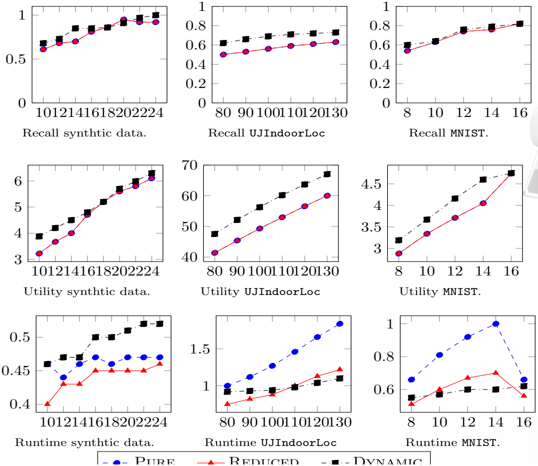
Predict (semantic) location, e.g. room number based on GPS. Use
$K = 80, \ldots, 130,\ \varepsilon = 0.1,\ \sigma = 1,\ k(e_i, e_j) = \exp\left(\frac{-||e_i - e_j||_2^2}{0.005}\right)$

MNIST

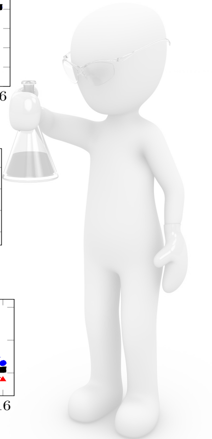
Handwritten digit recognition task. Use
$K = 8, \ldots, 16,\ \varepsilon = 0.1,\ \sigma = 1,\ k(e_i, e_j) = \exp\left(\frac{-||e_i - e_j||_2^2}{784}\right)$



Experiments Results



Recall synthtic data.
Recall UJIndoorLoc
Recall MNIST.
Utility synthtic data.
Utility UJIndoorLoc
Utility MNIST.
Runtime synthtic data.
Runtime UJIndoorLoc
Runtime MNIST.



Outlook

Question 1 Are summaries with IVM really expressive?
Yes! At least we have reasonable recall

Outlook

Question 1 Are summaries with IVM really expressive?
Yes! At least we have reasonable recall

Question 2 How well perform enhancements compared to vanilla?
Quite well! Computation decreased + utility and recall increased

Outlook

Question 1 Are summaries with IVM really expressive?
Yes! At least we have reasonable recall

Question 2 How well perform enhancements compared to vanilla?
Quite well! Computation decreased + utility and recall increased

Next to come
Better kernel functions?
Streaming with concept drift? ⇒ Maybe “forget“ items?
Use summaries for model learning?