Making Sense of Arduino-based Internet of Things

Matthieu PUIGT Clément DORFFER
Gilles DELMAIRE Gilles ROUSSEL

Univ. Littoral Côte d'Opale, EA 4491 – LISIC, F-62228 Calais, France
firstname.LASTNAME@lisic.univ-littoral.fr

April 2, 2016

Introduction (1)

Internet of Things (IoT)

*"One of the myths about the IoT is that companies have all the data they need, but their **real challenge is making sense of it**. In reality, [...] the quality of the data isn't always good enough, and it remains difficult to integrate multiple data sources."* – Chris Murphy

One key challenge in IoT

Trusting data

- Outlier /sensor fault detection
- Sensor calibration

Introduction (1)

Internet of Things (IoT)

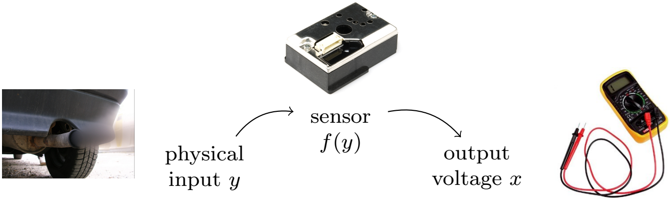
"One of the myths about the IoT is that companies have all the data they need, but their ***real challenge is making sense of it****. In reality, [...] the quality of the data isn't always good enough, and it remains difficult to integrate multiple data sources."* – Chris Murphy

One key challenge in IoT

Trusting data

- Outlier /sensor fault detection
- Sensor calibration (Ph.D. thesis of C. Dorffer)

Sensing process



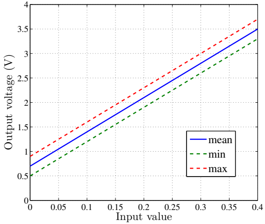
Introduction (2)

The "why" of calibration

Sensing process

physical input y → sensor $f(y)$ → output voltage x

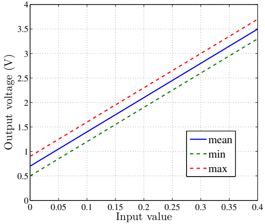
Typical manufacturer data

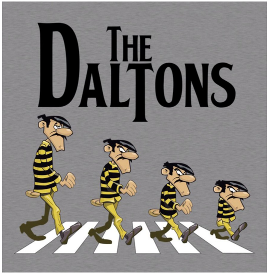


Sensing process

physical input y → sensor $f(y)$ → output voltage x

Typical manufacturer data





Introduction (2)

The "why" of calibration

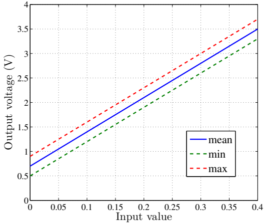
Sensing process

physical input y

sensor $f(y)$

output voltage x

Typical manufacturer data



- Imprecise reconstruction
- Error accumulation (multi-sensor)

  $\Rightarrow$ Calibration needed!

- Not always manually feasible (inaccessible sensors, too numerous...)

  $\Rightarrow$ Blind calibration

Introduction (3)

The "how" of blind sensor calibration

Calibration models $x(t) = f(y(t))$

- linear/affine (+ phase indeterminacy) $x(t) = \alpha_0 + \alpha_1 y(t)$
- polynomial $x(t) = \alpha_0 + \alpha_1 y(t) + \ldots + \alpha_N y(t)^N$
- multilinear $x(t) = \alpha_0 + \alpha_1 y(t) + \alpha_2 z(t)$

  + sensor drift $x(t) = \alpha_0(t) + \alpha_1(t) y(t)$

In-the-wild calibration methods

- Nullspace projection (Balzano *et al.* 2007/2014)
- Moments (Wang *et al.* 2008, Lee *et al.* 2014)
- Multi-hop (Saukh *et al.* 2015)
- Informed matrix factorization (Dorffer *et al.* 2015–)

Blind calibration of mobile sensors

- We are curently working on crowdsensing applications (with Inria Lille)
- Mobile sensors
- Blind calibration method based on the *rendezvous* assumption
- Brief outline of the method

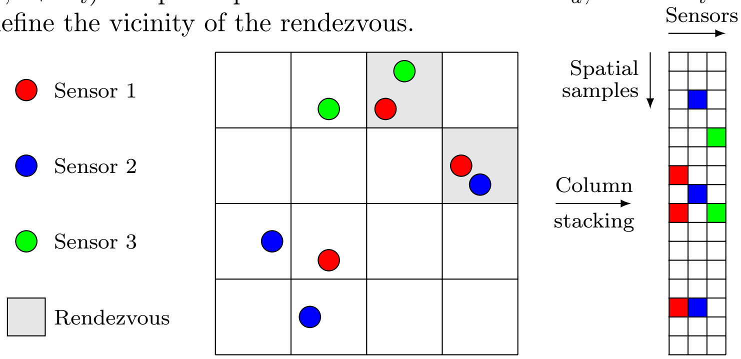
Definitions

Rendezvous

A **rendezvous** is a temporal and spatial vicinity between two sensors (Saukh *et al.* 2013).

Scene

A **scene** $\mathcal{S}$ is a discretized area observed during a time interval $[t, t + \Delta_t)$. A spatial pixel has a size lower than Δ_d, where Δ_t and Δ_d define the vicinity of the rendezvous.



Scene $\mathcal{S}$ Observed matrix X

Sensor response model

- Network = m mobile sensors
- Affine sensor response

$$x_{i,j} \approx (y_i \cdot \alpha_{1,j}) + \alpha_{0,j} \tag{1}$$

Global matrix form

$$\underbrace{\begin{bmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & & \vdots \\ x_{n,1} & \cdots & x_{n,m} \end{bmatrix}}_{X} \approx \underbrace{\begin{bmatrix} y_1 & 1 \\ \vdots & \vdots \\ y_n & 1 \end{bmatrix}}_{G} \cdot \underbrace{\begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,m} \\ \alpha_{0,1} & \alpha_{0,2} & \cdots & \alpha_{0,m} \end{bmatrix}}_{F}$$

In practice

Missing data in $X \Rightarrow W_{i,j} \triangleq \begin{cases} 0 & \text{if } x_{i,j} \text{ is not available,} \\ 1 & \text{otherwise} \end{cases}$

$$W \circ X \approx W \circ (G \cdot F)$$

- X, G and $F \geq 0$
- A known reference

$\Rightarrow$ $\forall i = 1,\ldots, n, \quad x_{i,m} = y_i$ (i.e., $\alpha_m = 1$, $\beta_m = 0$)

Matrix form with known reference and missing data

$$W \circ \underbrace{\begin{bmatrix} x_{1,1} & \cdots & x_{1,m-1} & y_1 \\ \vdots & & \vdots & \vdots \\ x_{n,1} & \cdots & x_{n,m} & y_n \end{bmatrix}}_{X} \approx W \circ \left(\underbrace{\begin{bmatrix} y_1 & 1 \\ \vdots & \vdots \\ y_n & 1 \end{bmatrix}}_{G} \cdot \underbrace{\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{m-1} & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_{m-1} & 0 \end{bmatrix}}_{F} \right)$$

Affine model + reference sensor $\rightarrow$ constrained values in G and F

Concepts of our earlier proposed method

- Example ($n = 4$, $m = 3$, $k = 2$ calibrated measurements)

$$W \circ \begin{bmatrix} x_{1,1} & x_{1,2} & y_1 \\ x_{2,1} & x_{2,2} & y_2 \\ x_{3,1} & x_{3,2} & y_3 \\ x_{4,1} & x_{4,2} & y_4 \end{bmatrix} \approx W \circ \left(\begin{bmatrix} y_1 & 1 \\ y_2 & 1 \\ y_3 & 1 \\ y_4 & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 & \alpha_2 & 1 \\ \beta_1 & \beta_2 & 0 \end{bmatrix} \right)$$

- Only update the free parts of G and F
- Works well if the sensor network is dense enough ($\simeq$ matrix completion assumptions), e.g.,

$$W \circ X = \begin{bmatrix} x_{1,1} & x_{2,2} & - \\ x_{2,1} & - & y_2 \\ - & x_{3,2} & y_3 \\ - & x_{4,2} & - \end{bmatrix}$$

- What if the sensor network is not dense enough?

$$W \circ X = \begin{bmatrix} x_{1,1} & - & - \\ - & - & y_2 \\ - & - & y_3 \\ - & x_{4,2} & - \end{bmatrix}$$

- What if the sensor network is not dense enough?

$$W \circ X = \begin{bmatrix} x_{1,1} & - & - \\ - & - & y_2 \\ - & - & y_3 \\ - & x_{4,2} & - \end{bmatrix}$$

- Sparse decomposition of the scene, i.e., $y \approx \tilde{y} = \mathcal{D} \cdot a$ with $\mathcal{D} \in \mathbb{R}^{n,l}$ known dictionary, a **sparse** vector of contributions

- What if the sensor network is not dense enough?

$$W \circ X = \begin{bmatrix} x_{1,1} & - & \tilde{y}_1 \\ - & - & y_2 \\ - & - & y_3 \\ - & x_{4,2} & \tilde{y}_4 \end{bmatrix} \text{with } \tilde{y} = \mathcal{D} \cdot a$$

- Sparse decomposition of the scene, i.e., $y \approx \tilde{y} = \mathcal{D} \cdot a$ with $\mathcal{D} \in \mathbb{R}^{n,l}$ known dictionary, a **sparse** vector of contributions

A new informed NMF model

Novel optimization problem

- What if the sensor network is not dense enough?

$$W \circ X = \begin{bmatrix} x_{1,1} & - & \tilde{y}_1 \\ - & - & y_2 \\ - & - & y_3 \\ - & x_{4,2} & \tilde{y}_4 \end{bmatrix} \text{ with } \tilde{y} = \mathcal{D} \cdot a$$

- Sparse decomposition of the scene, i.e., $y \approx \tilde{y} = \mathcal{D} \cdot a$ with $\mathcal{D} \in \mathbb{R}^{n,l}$ known dictionary, a **sparse** vector of contributions

➩ **“Virtual” rendezvous**

A new informed NMF model

Novel optimization problem

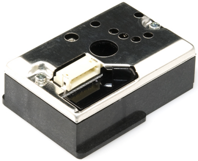
- What if the sensor network is not dense enough?

$$W \circ X = \begin{bmatrix} x_{1,1} & - & \tilde{y}_1 \\ - & - & y_2 \\ - & - & y_3 \\ - & x_{4,2} & \tilde{y}_4 \end{bmatrix} \text{ with } \tilde{y} = \mathcal{D} \cdot a$$

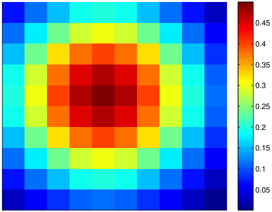
- Sparse decomposition of the scene, i.e., $y \approx \tilde{y} = \mathcal{D} \cdot a$ with $\mathcal{D} \in \mathbb{R}^{n,l}$ known dictionary, a **sparse** vector of contributions

➯ **“Virtual” rendezvous**

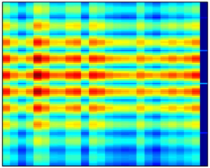
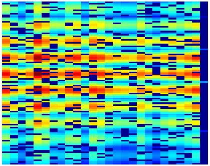
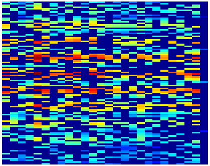
- New approach where we estimate/update G, F, and a



- 10×10 discretized area ($n = 100$)
- Values $0 < y_i \leq 0.5$ mg/m^3
- 62 atoms dictionnary
- Decomposition of y with 2 atoms

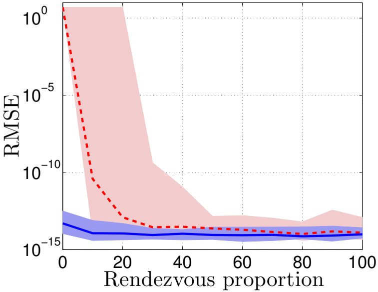


- $m = 26$ mobile dust sensors
- Parameters (gain+offset) chosen according to Gaussian distribution around theoretical parameters.
  $\forall j = 1 \cdots m-1, \quad 3.5 < \alpha_j < 6.5$ and $0 < \beta_j < 1.5$
- X is a 100×26 data matrix with
  - $k = 4$ reference measurements
  - l mobile measurements
- Tests w.r.t. number of rendezvous between calibrated and uncalibrated sensors, number of missing values, and noise

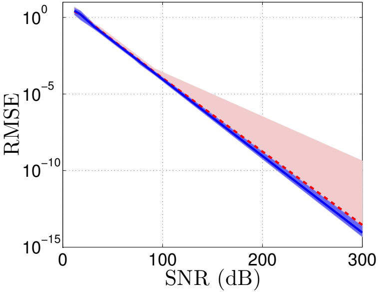
RMSE vs Rendezvous proportion

- 90% missing values
- Noiseless data
- 25 random runs / condition
- 2.e5 iterations
- Perf. criterion: median and envelope of RMSEs
  - In blue: the proposed approach
  - In red: our previous method



RMSE vs SNR

- 30% Rendezvous
- 90% Missing data
- 25 random runs / condition
- $2.e5$ iterations
- Perf. criterion: same as above



Accuracy of reconstruction of the Scene

MPMP Mettre la figure du poster Liban

Conclusion

- Some expertise in blind mobile sensor calibration
- Relaxed rendezvous provide a way to also solve blind fixed sensor calibration (up to some new development)
- We also have experience in informed matrix factorization robust to outliers (previous Ph.D. thesis)

What can ten be done?

- Optimization of the sensor network placement (talk by G. Roussel)