DATA-SELECTIVE LMS-NEWTON AND LMS-QUASI-NEWTON ALGORITHMS

Christos Tsinos* and Paulo S. R. Diniz**

*University of Luxembourg

**Federal University of Rio de Janeiro

ICASSP 2019

14/05/2019

Outline

- Introduction
- System Model
- LMS-Newton and LMS-Quasi-Newton Algorithms
- Data-Selective Approaches
- Simulations

Introduction (1/3)

SnT securityandtrust.lu

- In the era of big data, the processing will demand huge computational load if an effective strategy is not followed
- Data-selective processing
  - Process only the innovative new data
  - May avoid outliers
  - Performance close to the one of the non-data-selective counterparts
  - Reduced computational burden since only a very small portion of the data is processed

uni.lu UNIVERSITÉ DU LUXEMBOURG

Introduction (2/3)

- This paper develops data-selective versions of
  - LMS-Newton (LMSN)
  - LMS-Quasi-Newton (LMSQN)
- LMSN/LMSQN are powerful alternatives to the classical LMS
  - Higher Complexity
  - Better Performance in several cases (e.g. when the spread of the eigenvalues of the input-signal correlation matrix is large)
  - Some versions of LMSQN appear to be very robust to quantization errors compared to algorithms of similar complexity/performance, i.e., RLS.

Introduction (3/3)

- The data are classified via two thresholds as
  - Non-innovative
  - Innovative
  - Outliers
- The thresholds are tuned based on a prescribed probability of update
- The latter probability is connected to the Mean Square Error (MSE) of the algorithms
- The performance is evaluated via simulations on synthetic and real world data

System Model (1/2)

securityandtrust.lu

- Linear System Identification Problem
- Input-Output

$$d(k) = \mathbf{w}_o^T \mathbf{x}(k) + n(k)$$

- $\mathbf{w}_o \in \mathbb{R}^{L+1}$ is the unknown system
- $\mathbf{x}(k) = [x(k)\; x(k-1)\; \ldots\; x(k-L+1)]^T$ is the input signal
- $n(k)$ is a Gaussian noise sample of variance σ_n^2
- A filtering algorithm generates an output signal estimation via $\mathbf{w}^T(k)\mathbf{x}(k)$

UNIVERSITÉ DU LUXEMBOURG

System Model (2/2)

- Error Estimation Sequence for $k = 0, 1, \ldots, \infty$

$$e(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k)$$

- Mean Square Error

$$\begin{aligned} \xi(k) =& \sigma_n^2 + \mathbb{E}\{\Delta\mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\Delta\mathbf{w}(k)\} \\ =& \sigma_n^2 + \xi_{exc}(k), \end{aligned}$$

- $\Delta\mathbf{w}(k) = \mathbf{w}(k) - \mathbf{w}_o$
- The MSE formula is used to prescribe the desired probability of update

LMSN/LMSQN Algorithms (1/3)

- The aim is to minimize the cost function

$$J(\mathbf{w}(k)) = \frac{1}{2}|e(k)|^2$$

- Update Step

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \frac{\mu}{\mathbf{x}^T(k)\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)}\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)\tilde{e}(k)$$

- μ is a step-size parameter
- $\tilde{e}(k) = d(k) - \mathbf{w}^T(k-1)\mathbf{x}(k)$ is the a priori estimation error
- $\hat{\mathbf{R}}(k)$ is the estimation of $\mathbf{R} = \mathbb{E}\{\mathbf{x}(k)\mathbf{x}^T(k)\}$

LMSN/LMSQN Algorithms (2/3)

- LMSN and LMSQN differ on how $\hat{\mathbf{R}}^{-1}(k)$ is estimated
- LMSN estimation is based on a Robbins-Monro procedure

$$\hat{\mathbf{R}}^{-1}(k) = \frac{1}{1-\alpha}\left\{\hat{\mathbf{R}}^{-1}(k-1) - \frac{\hat{\mathbf{R}}^{-1}(k-1)\mathbf{x}(k)\mathbf{x}^T(k)\hat{\mathbf{R}}^{-1}(k-1)}{\frac{1-a}{a} + \mathbf{x}(k)^T\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)}\right\}$$

- α is a step-size parameter

LMSN/LMSQN Algorithms (3/3)

SnT securityandtrust.lu

- For the LMSQN the estimation is given by

$$\hat{\mathbf{R}}^{-1}(k) = \frac{1}{1-\alpha}\left\{\hat{\mathbf{R}}^{-1}(k-1) + \left(\frac{\mu}{2\mathbf{x}(k)^T\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)} - 1\right)\right.$$
$$\left. \times \frac{\hat{\mathbf{R}}^{-1}(k-1)\mathbf{x}(k)\mathbf{x}^T(k)\hat{\mathbf{R}}^{-1}(k-1)}{\mathbf{x}(k)^T\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)}\right\}.$$

P. S. R. Diniz, M. L. R. de Campos, and A. Antoniou, "Analysis of LMS-Newton adaptive filtering algorithms with variable convergence factor," IEEE Trans. Signal Process., vol. 43, no.3, pp. 617–627, March 1995.

M. L. R. De Campos and A. Antoniou, "A new quasi-Newton adaptive filtering algorithm," IEEE Trans. Circuits Syst. II. Analog Digit. Signal Process., vol. 44, no. 11, pp. 924–934, Nov. 1997.

Data-Selective Approaches (1/5)

- New data are classified as innovative if $|e(k)|^2$ is greater than a scaled noise power level $\tau(k)\sigma_n^2$
- If $|e(k)|^2$ is greater than $\tau_{max}\sigma_n^2$, an outlier is identified and no update is performed
- Equivalent cost function

$$J'(\mathbf{w}(k)) = \begin{cases} \frac{1}{2}|e(k)|^2, \text{ if } \sqrt{\tau(k)} \leq \frac{|e(k)|}{\sigma_n} < \sqrt{\tau_{max}} \\ 0, \text{ otherwise.} \end{cases}$$

Data-Selective Approaches (2/5)

securityandtrust.lu

- Update for the data-selective approach

$$\mathbf{w}(k) = \begin{cases} \mathbf{w}(k-1) + \mu \frac{\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)\tilde{e}(k)}{\mathbf{x}^T(k)\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)}, & \sqrt{\tau(k)} \leq \frac{|e(k)|}{\sigma_n} < \sqrt{\tau_{max}} \\ \mathbf{w}(k-1), & \text{otherwise.} \end{cases}$$

- The data-selective strategy may be adopted for the update of $\hat{\mathbf{R}}^{-1}$, as well.
- Desired probability of update

$$P_{up}(k) = P\left\{\frac{|e(k)|}{\sigma_n} > \sqrt{\tau(k)}\right\} - P\left\{\frac{|e(k)|}{\sigma_n} > \sqrt{\tau_{max}}\right\}$$

UNIVERSITÉ DU LUXEMBOURG

Data-Selective Approaches (3/5)

- Under the assumption of white Gaussian input signals, at the steady state we have

$$P_{up} = 2Q\left(\frac{\sigma_n\sqrt{\tau}}{\sigma_e}\right) - 2Q\left(\frac{\sigma_n\sqrt{\tau_{max}}}{\sigma_e}\right)$$

- $Q(\cdot)$ is the complementary Gaussian cumulative distribution function
- σ_e^2 is the error signal variance
- Index k is dropped under the assumption of stationarity

Data-Selective Approaches (4/5)

securityandtrust.lu

- *Proposition: The excess mean square error at the steady-state can be approximated by*

$$\xi_{exc}(\infty) = \frac{\mu P_{up}}{2 - \mu P_{up}} \sigma_n^2$$

- If no outliers are presented, the threshold is $\sqrt{\tau} = \sqrt{1 + \beta} Q^{-1}(0.5 P_{up})$, $\beta = \frac{\mu P_{up}}{2 - \mu P_{up}}$
- For the case of outliers, some prior information of the signal sources and supporting circuitry, is needed for deriving the thresholds

M.O.K. Mendonc¸a, J. O. Ferreira, C. G. Tsinos, P.S.R. Diniz, and T. N. Ferreira, "On fast converging data-selective adaptive filtering," Algorithms, vol. 12, no. 1, pp. 4, 2019.

UNIVERSITÉ DU LUXEMBOURG

Data-Selective Approaches (5/5)

securityandtrust.lu

Algorithm 1 Data-selective LMSN and LMSQN Algorithms

Inputs: $0 < \mu \leq 1$, $0 < \alpha \leq 1$ (for LMSN), γ small positive value, P_{up} and τ_{max}
Initialize $\mathbf{w}(0) = \mathbf{0}_{L+1}$ and $\hat{\mathbf{R}}^{-1}(0) = \gamma \mathbf{I}_{L+1}$
Set $\beta = \frac{\mu P_{up}}{2-\mu P_{up}}$
Calculate τ from (13), if outliers are present or from (16), otherwise
for $k = 1, 2, \ldots$ **do**
  Acquire $\mathbf{x}(k)$ and $d(k)$
  $e(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k)$
  if $\sqrt{\tau}\sigma_n \leq |e(k)| \leq \sqrt{\tau_{max}}\sigma_n$ **then**
    $\mathbf{t}(k) \leftarrow \hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)$
    $\psi(k) \leftarrow \mathbf{x}^T(k)\mathbf{t}(k)$
    $\mathbf{w}(k+1) \leftarrow \mathbf{w}(k) + \mu \frac{\mathbf{t}(k)e(k)}{\psi(k)}$
    $\hat{\mathbf{R}}^{-1}(k+1) \leftarrow \frac{1}{1-a}\left[\hat{\mathbf{R}}^{-1}(k) - \frac{\mathbf{t}(k)\mathbf{t}^T(k)}{\frac{1-\alpha}{\alpha}+\psi(k)}\right]$, for LMSN
    $\hat{\mathbf{R}}^{-1}(k+1) \leftarrow \hat{\mathbf{R}}^{-1}(k) + \frac{\frac{\mu}{2\psi(k)}-1}{\psi(k)}\mathbf{t}(k)\mathbf{t}^T(k)$, for LMSQN
  else if $|e(k)| \leq \sqrt{\tau}\sigma_n$ **then**
    $\mathbf{w}(k+1) \leftarrow \mathbf{w}(k)$
  else if $|e(k)| \geq \sqrt{\tau_{max}}\sigma_n$ **then**
    $\mathbf{w}(k+1) \leftarrow \mathbf{w}(k)$, $e(k) = 0$, $d(k) = 0$
  end if
end for

UNIVERSITÉ DU LUXEMBOURG

Simulations (1/5)

- System identification problem

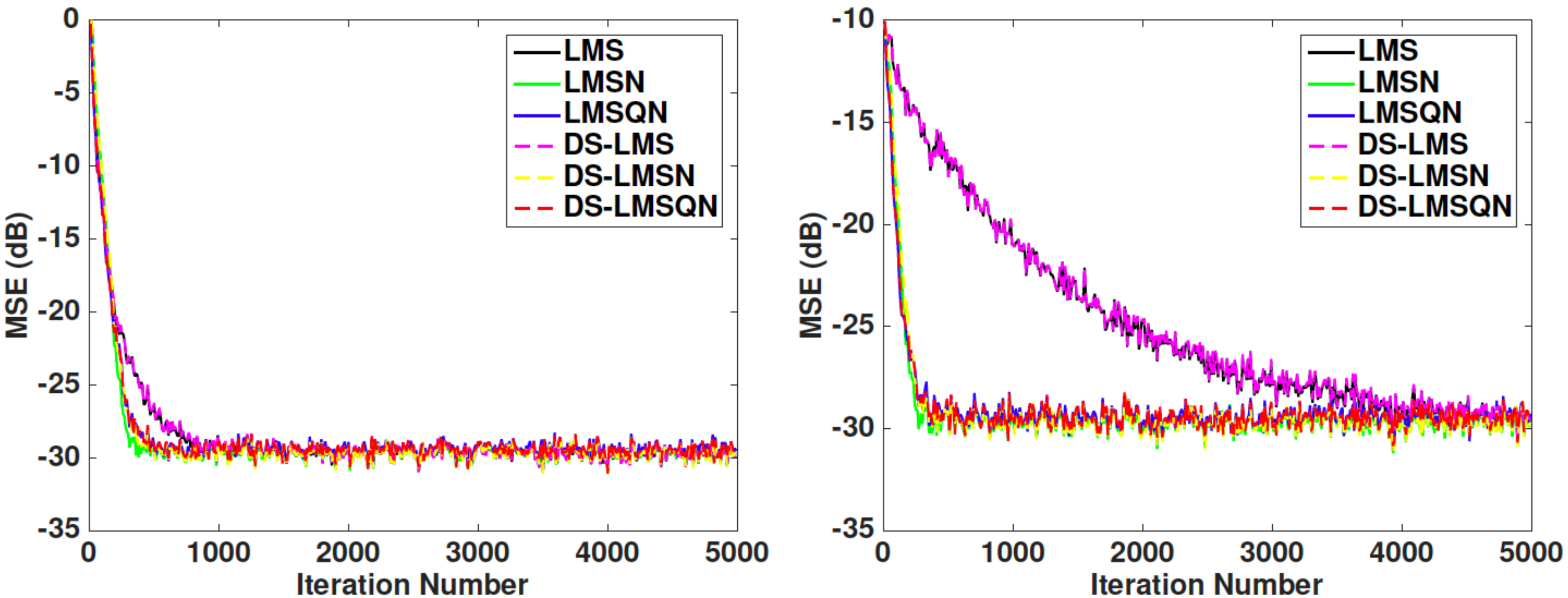
$$[0.1010\ 0.3030\ 0\ -0.2020\ -0.4040$$
$$-0.7071\ -0.4040\ -0.2020].$$

- Input signals

$$x(k) = 0.88x(k-1) + n_1(k),$$
$$x(k) = -0.55x(k-1) - 1.221x(k-2) - 0.49955x(k-3)$$
$$-0.4536x(k-1) + n_2(k),$$

- $n_1(k)$ and $n_2(k)$ are uncorrelated Gaussian noise variables

Simulations (2/5)

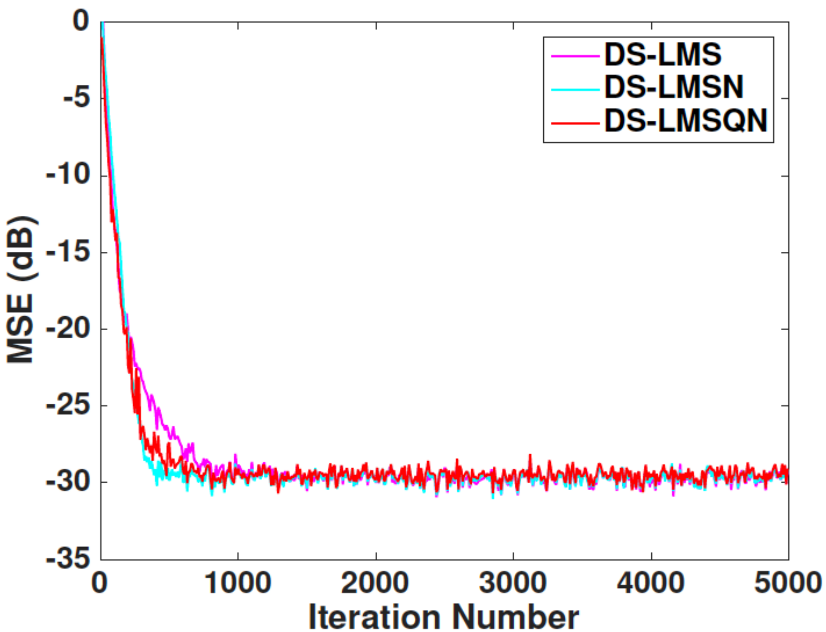


(a) AR(1) (b) AR(4)

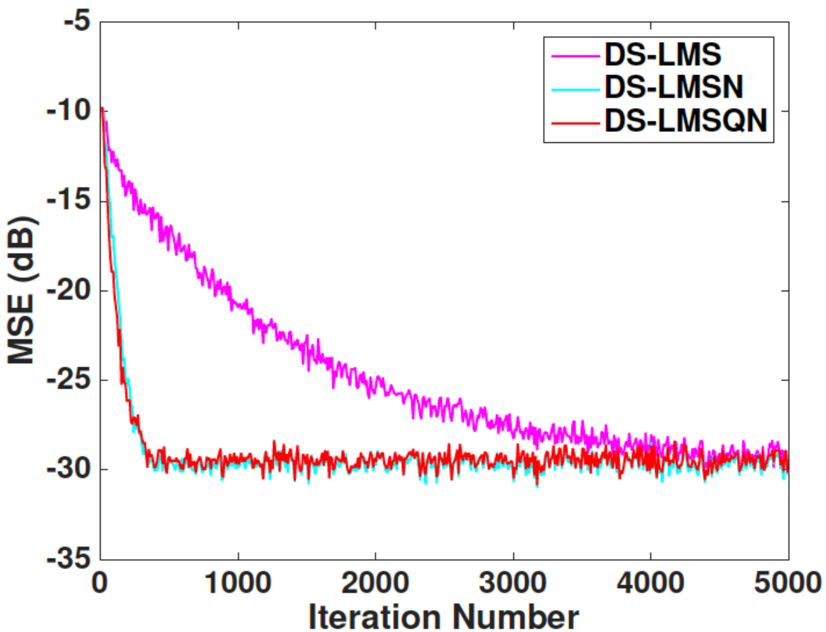
MSE Learning Curves - No Outliers - $P_{up} = 0.4$

Simulations (3/5)

securityandtrust.lu



(c) AR(1) + outliers

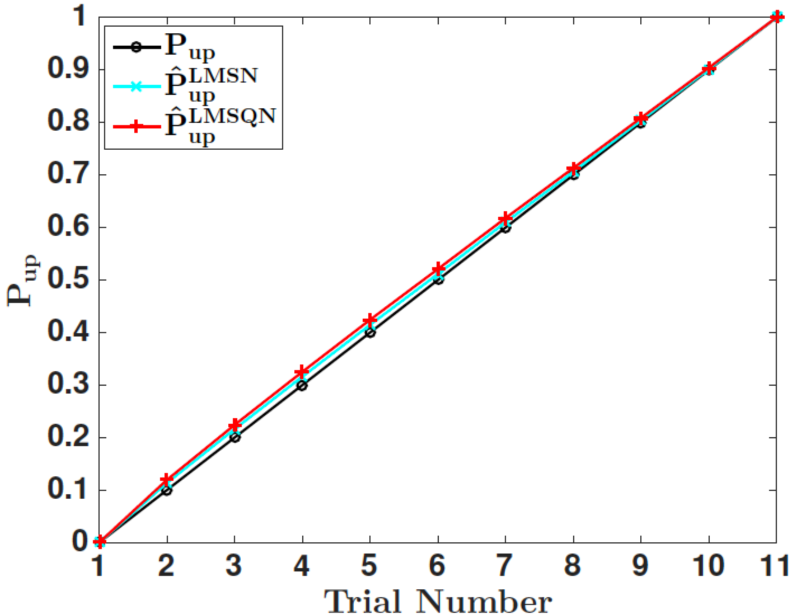


(d) AR(4) + outliers

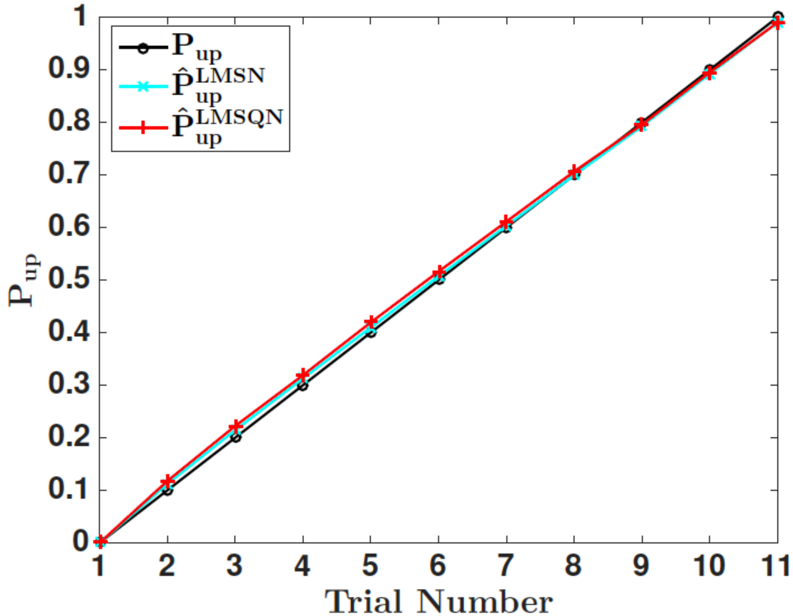
MSE Learning Curves - Outliers - $P_{up} = 0.4$

UNIVERSITÉ DU LUXEMBOURG

Simulations (4/5)



(a) No outliers

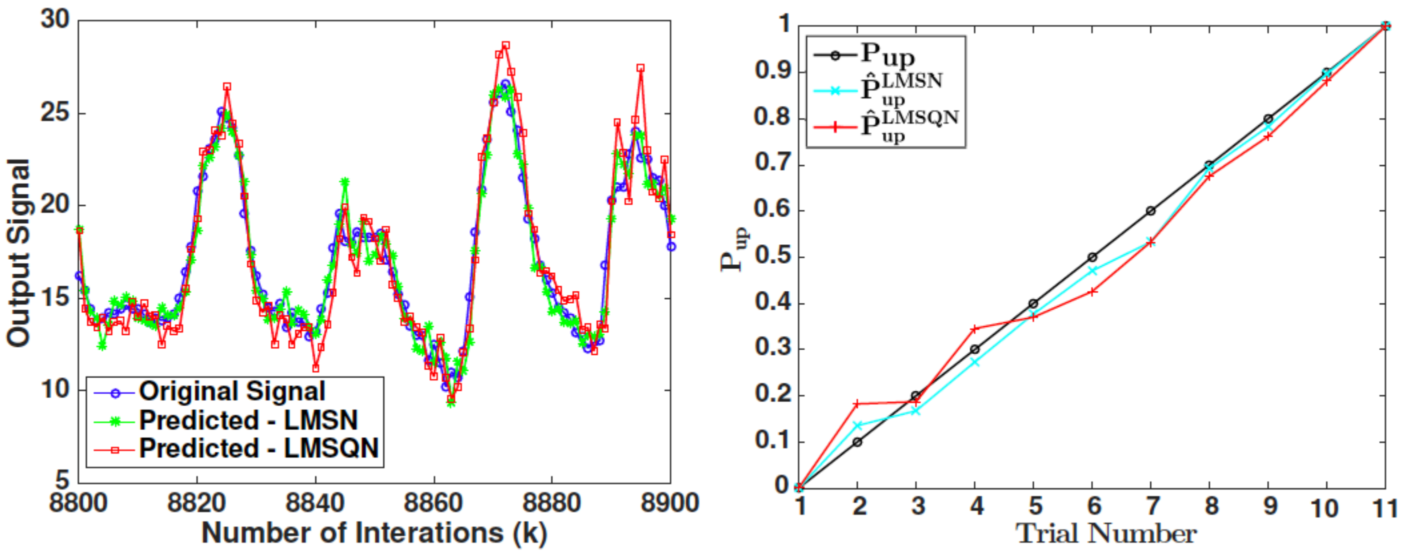


(b) Outliers

Fig. 2. Comparison between prescribed P_{up} and the achieved $\hat{P}_{up}^{LMSN}$ and $\hat{P}_{up}^{LMSQN}$ by the data-selective LMSN and LMSQN algorithms.

Simulations (5/5)

- Temperature prediction on a data-set provided by University of California at Irvine
- The prediction error variance σ_e^2 is derived according to the setup - $P_{up} = 0.4$



(a) (b)

UC Irvine, "Air quality data set, machine learning repository, [online]," https://archive.ics.uci.edu/ml/datasets/Air+quality, Accessed: 2017-09.

Conclusion

securityandtrust.lu

- Data Selective LMSN and LMSQN algorithms were developed
- Computational overhead reduction
- Outlier Exclusion from the learning process
- Performance evaluated via simulations on synthetic and real world data
- Extensions on distributed adaptive filtering under development

Thank you for your attention