Recovery of Missing Sensor Data by Reconstructing Time-varying Graph Signals

Anindya Mondal Mayukhmali Das Aditi Chatterjee Palaniandavar Venkateswaran

Department of Electronics and Telecommunication Engineering
Jadavpur University, India

EUSIPCO 2022

Session: *Signal Processing over Graphs and Networks*

August 30, 2022 - Belgrade, Serbia

Presentation overview

1. Motivation & Challenges
2. Primary Contributions
3. Proposed Algorithm
4. Experimental Framework

EUSIPCO BELGRADE/SERBIA 29 AUG – 2 SEP 2022

Table of Contents

Motivation and Challenges

Wireless Sensor Networks (WSNs) are among the most promising technologies of the current era. Primary reasons for their widespread application include:

- Low cost

Motivation and Challenges

Wireless Sensor Networks (WSNs) are among the most promising technologies of the current era. Primary reasons for their widespread application include:

- Low cost
- Small size

Motivation and Challenges

Wireless Sensor Networks (WSNs) are among the most promising technologies of the current era. Primary reasons for their widespread application include:

- Low cost
- Small size
- Low power consumption

Motivation and Challenges

Wireless Sensor Networks (WSNs) are among the most promising technologies of the current era. Primary reasons for their widespread application include:

- Low cost
- Small size
- Low power consumption
- Easy to deploy

Often these sensors are **deployed in remote locations**.

Often these sensors are **deployed in remote locations**.

Sensor nodes may get **damaged** due to **environmental/ human intervention**.

Often these sensors are **deployed in remote locations**.

Sensor nodes may get **damaged** due to **environmental/ human intervention**.

Replacing the malfunctioning/ inactive sensors becomes a **tedious task**.

Often these sensors are **deployed in remote locations**.

Sensor nodes may get **damaged** due to **environmental/ human intervention**.

Replacing the malfunctioning/ inactive sensors becomes a **tedious task**.

Accurately reconstructing the missing values is the only way forward.

State-of-the-art recovery approaches were based on:

- k-Nearest-Neighbours (kNN)

State-of-the-art recovery approaches were based on:

- k-Nearest-Neighbours (kNN)
- Expectation-Maximization (E-M)

State-of-the-art recovery approaches were based on:

- k-Nearest-Neighbours (kNN)
- Expectation-Maximization (E-M)
- Low-rank Matrix Completion (LRMC)

State-of-the-art recovery approaches were based on:

- k-Nearest-Neighbours (kNN)
- Expectation-Maximization (E-M)
- Low-rank Matrix Completion (LRMC)
- Probabilistic Matrix Factorization (PMF)

Most of these methods give sub-optimal performances when a **sizable amount of data is unavailable**.

Most of these methods give sub-optimal performances when a **sizable amount of data is unavailable**.

They failed to **properly** utilize the embedded spatio-temporal dependencies in the sensor data.

Most of these methods give sub-optimal performances when a **sizable amount of data is unavailable**.

They failed to **properly** utilize the embedded spatio-temporal dependencies in the sensor data.

The proposed method improves upon them to a great extent.

Table of Contents

Primary Contributions

Use of graph-based methods has increased in various domains like computer vision, biological networks, social networks, infectious disease prediction and many others.

Primary Contributions

Use of graph-based methods has increased in various domains like computer vision, biological networks, social networks, infectious disease prediction and many others.

Primary reason behind this is the ability of graphs to capture meaningful information from data lying on irregular and complex structures.

Primary Contributions

Use of graph-based methods has increased in various domains like computer vision, biological networks, social networks, infectious disease prediction and many others.

Primary reason behind this is the ability of graphs to capture meaningful information from data lying on irregular and complex structures.

Graph Signal Processing (GSP) is one such graph-based technique. Sampling and reconstruction of graph signals in GSP is useful for various real-life applications.

Primary Contributions

Use of graph-based methods has increased in various domains like computer vision, biological networks, social networks, infectious disease prediction and many others.

Primary reason behind this is the ability of graphs to capture meaningful information from data lying on irregular and complex structures.

Graph Signal Processing (GSP) is one such graph-based technique. Sampling and reconstruction of graph signals in GSP is useful for various real-life applications.

In our work, the graph vertices represent the relative position of the sensor nodes and the vertex-indexed signals represent the sensor attributes.

To summarize our contributions:

- Introducing the concept of **reconstructing time-varying graph signals** to **recover missing data** in wireless sensor networks.

To summarize our contributions:

- This work is inspired by the 2020 MLSP paper by Giraldo *et al.*

To summarize our contributions:

- There they **predict the number of new COVID-19 cases** by extending the **Sobolev norm** defined in GSP for **time-varying graph signals**.

To summarize our contributions:

- There they **predict the number of new COVID-19 cases** by extending the **Sobolev norm** defined in GSP for **time-varying graph signals**.
- Here we utilize the Sobolev method for the recovery purpose

To summarize our contributions:

- There they **predict the number of new COVID-19 cases** by extending the **Sobolev norm** defined in GSP for **time-varying graph signals**.
- Here we utilize the Sobolev method for the recovery purpose
- The Sobolev algorithms **shows significant performance improvement over SoTA**.

To summarize our contributions:

- There they **predict the number of new COVID-19 cases** by extending the **Sobolev norm** defined in GSP for **time-varying graph signals**.
- Here we utilize the Sobolev method for the recovery purpose
- The Sobolev algorithms **shows significant performance improvement over SoTA**.
- We validate our claim by performing experiments on **several publicly available datasets** collected from diverse environments (indoor and outdoor).

Table of Contents

Proposed Algorithm

Graph Construction

$G = (\nu, \mathcal{E}, \mathbf{W})$ is a graph with ν as the set of vertices, $\mathcal{E}$ as the set of edges and $\mathbf{W}$ as the weighted adjacency matrix.

$\mathbf{D}$ is the degree matrix and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the unnormalized Laplacian matrix of G.

$x : \nu \to \mathbb{R}$ is a vertex-indexed graph signal.

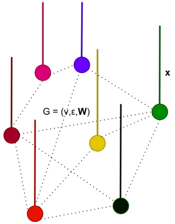

Figure: Graph and graph signals.

Sampling and Reconstruction of Graph Signals

$\boldsymbol{\theta}$ is a sampling operator.

$\mathbf{x}(S) = \boldsymbol{\theta}\mathbf{x}$, S is the sequence of sampled indices.

$\boldsymbol{\phi}$ is the interpolation operator.

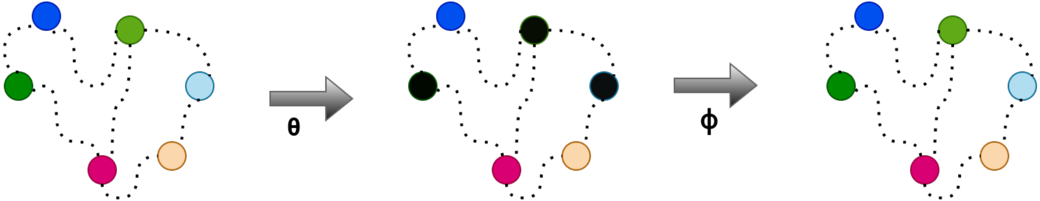

Figure: Graph signals sampling and reconstruction.

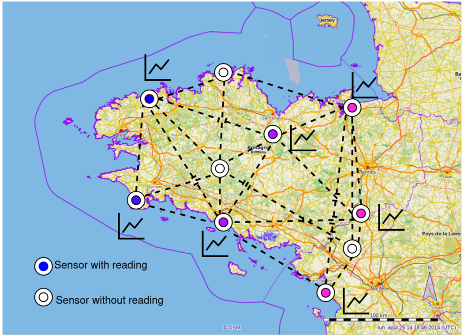

Figure: A sensor network.

- We model the missing sensor data recovery problem as the **reconstruction of time-varying graph signals**.

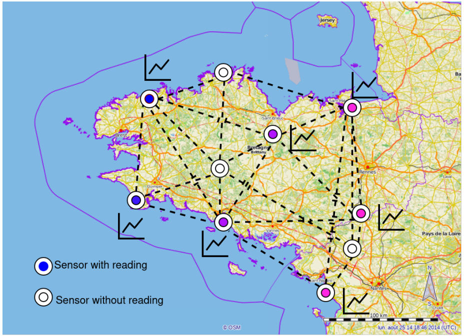

Figure: A sensor network.

- We model the missing sensor data recovery problem as the **reconstruction of time-varying graph signals**.
- We perform the reconstruction by **minimizing the Sobolev norm of time-varying graph signals**.

NOTE:

- We have made a prior assumption that the graph signals are smooth in space and time.

NOTE:

- We have made a prior assumption that the graph signals are smooth in space and time.
- This means that nearby sensors have similar readings and changes in readings are not abrupt with time.

Let $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_M}]^T$ be a time-varying graph signal, where $\mathbf{x}_t$ denotes the signal at time t $(1 < t < M)$ in G. Here each row of $\mathbf{X}$ represents a time-series on the corresponding vertex.

Let $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_M}]^T$ be a time-varying graph signal, where $\mathbf{x}_t$ denotes the signal at time t $(1 < t < M)$ in G. Here each row of $\mathbf{X}$ represents a time-series on the corresponding vertex.

According to Qiu *et al.* (2017), we define the smoothness function as:

$$S_2(\mathbf{X}) = \sum_{t=1}^{M} S_2(\mathbf{x_t}) = \mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{L}\mathbf{X}\right).$$

EUROPEAN SIGNAL PROCESSING CONFERENCE 30th EUSIPCO BELGRADE/SERBIA 29 AUG - 2 SEP 2022

Let $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_M}]^T$ be a time-varying graph signal, where $\mathbf{x}_t$ denotes the signal at time t $(1 < t < M)$ in G. Here each row of $\mathbf{X}$ represents a time-series on the corresponding vertex.

According to Qiu *et al.* (2017), we define the smoothness function as:

$$S_2(\mathbf{X}) = \sum_{t=1}^{M} S_2(\mathbf{x_t}) = \mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{L}\mathbf{X}\right).$$

To include the temporal information, we also define the temporal difference operator as:

$$\mathbf{D}_h = \begin{bmatrix} -1 & & & \\ 1 & -1 & & \\ & 1 & \ddots & \\ & & \ddots & -1 \\ & & & 1 \end{bmatrix} \tag{1}$$

and the temporal difference signal as:

$$\mathbf{X}\mathbf{D}_h = [\mathbf{x}_2 - \mathbf{x}_1,\ \mathbf{x}_3 - \mathbf{x}_2,\ \ldots,\ \mathbf{x}_M - \mathbf{x}_{M-1}]$$

Now we define the sampling matrix for the whole time-varying graph signal as:

$$\mathbf{J}(i,\ j) = \begin{cases} 1 \textbf{ if } i \in S_t, \\ 0 \textbf{ if } i \notin S_t \end{cases} \tag{2}$$

Now we define the sampling matrix for the whole time-varying graph signal as:

$$\mathbf{J}(i,\ j) = \begin{cases} 1 \ \textbf{if} \ i \in S_t, \\ 0 \ \textbf{if} \ i \notin S_t \end{cases} \tag{2}$$

As per Giraldo *et al.* (2020), we define Sobolev norm for time-varying graph signals as follows:

$$\|\mathbf{X}\|_{\beta,\epsilon} = \sum_{i=1}^{M} \mathbf{x}_i^{\mathbf{T}}(\mathbf{L} + \epsilon\mathbf{I})^{\beta}\mathbf{x_i} = \mathbf{tr}(\mathbf{X}^{\mathbf{T}}(\mathbf{L} + \epsilon\mathbf{I})^{\beta}\mathbf{X}), \tag{3}$$

L being a symmetric matrix.

Now we define the sampling matrix for the whole time-varying graph signal as:

$$\mathbf{J}(i,\ j) = \begin{cases} 1 \ \textbf{if} \ i \in S_t, \\ 0 \ \textbf{if} \ i \notin S_t \end{cases} \tag{2}$$

As per Giraldo *et al.* (2020), we define Sobolev norm for time-varying graph signals as follows:

$$\|\mathbf{X}\|_{\beta,\epsilon} = \sum_{i=1}^{M} \mathbf{x}_i^{\mathbf{T}}(\mathbf{L}+\epsilon\mathbf{I})^{\beta}\mathbf{x_i} = \mathbf{tr}(\mathbf{X}^{\mathbf{T}}(\mathbf{L}+\epsilon\mathbf{I})^{\beta}\mathbf{X}), \tag{3}$$

$\mathbf{L}$ being a symmetric matrix.

Finally, we formulate the Sobolev reconstruction problem for time-varying graph signals as:

$$\min_{\overline{\mathbf{X}}} \frac{1}{2}\|\mathbf{J}\circ\overline{\mathbf{X}} - \mathbf{Y}\|_F^2 + \frac{\gamma}{2}\mathbf{tr}((\overline{\mathbf{X}}\mathbf{D}_h)^{\mathbf{T}}(\mathbf{L}+\epsilon\mathbf{I})^{\beta}\overline{\mathbf{X}}\mathbf{D}_h). \tag{4}$$

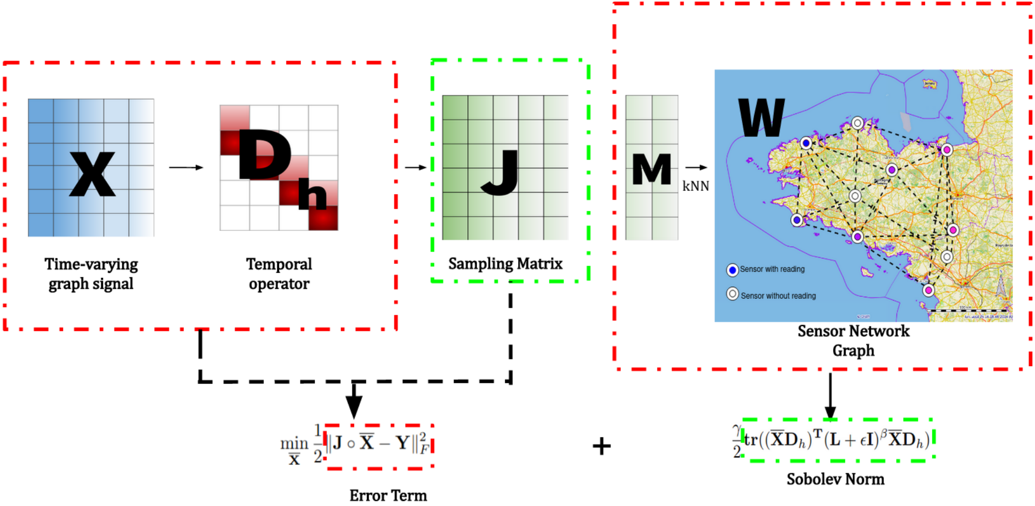

Figure: Schematic diagram.

Table of Contents

Experimental Framework

We evaluate the performance of the proposed method on the following datasets:

- **Moléne Dataset**: An open-access dataset of hourly weather observations in Brittany, France [Girault (2015)].
- **Intel Lab Dataset**: Data collected from 54 sensors deployed in the Intel Berkeley Research Laboratory [Madden (2004)].

We compare the method against SoTA recovery approaches (**kNN** [Kiani et al. (2017)], **EM** [Zhang et al. (2012)], **LRMC** [López-Valcarce et al. (2019)] and **PMF** [Fekade et al (2017)]).

We **randomly remove** some data points for each dataset and then **reconstruct the missing values** using the proposed framework.

Thereafter we compare the missing values with their corresponding original values from the dataset using RMSE and MAE.

	Sampling Density	kNN [2], [3]		EM [4]		LRMC [5], [6]		PMF [7]		Proposed	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Molene Dataset	0.1	8.05	6.75	7.34	5.68	3.47	2.35	4.75	2.56	**2.40**	**1.65**
	0.3	5.45	4.68	5.20	3.93	2.82	1.35	3.42	1.64	**1.60**	**1.02**
	0.5	4.20	3.04	4.14	2.95	1.96	1.02	2.45	1.09	**1.20**	**0.74**
	0.7	3.40	2.13	3.19	1.82	1.02	0.81	1.22	0.95	**0.97**	**0.65**
Intel Dataset	0.1	9.70	5.34	7.78	5.63	4.68	2.94	5.34	3.98	**3.04**	**1.50**
	0.3	6.30	3.52	6.45	4.96	3.40	2.15	3.95	2.37	**2.01**	**0.97**
	0.5	4.60	2.98	5.17	3.81	2.06	1.21	2.75	1.65	**1.57**	**0.70**
	0.7	3.75	1.78	4.63	2.57	1.45	0.75	1.97	0.99	**1.30**	**0.55**

Figure: Quantitative comparison of SoTA methods vs proposed method. Best results are in **bold**

Sampling density means the **amount of valid/ available information**, e.g. sampling density of 0.1 means that we have 10% of the data available and we need to reconstruct 90% of the data.

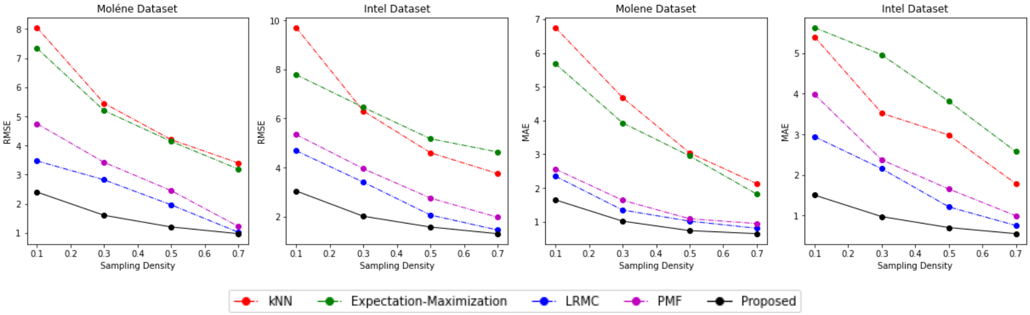

Figure: Performance of various SoTA algorithms in recovering missing sensor data.

This work applies the concept of time-varying graph signals reconstruction to recover missing data in sensor networks.

It leverages the concept of a recently introduced method based on minimization of the Sobolev norm of time-varying graph signals.

This method surpasses multiple previous approaches in the reconstruction accuracy, especially during the unavailability of a big chunk of data.

Future work may include:

- Extending the method to non-smooth situations, e.g. when nearby localities have drastically different temperatures or changes in temperature are abrupt with time.
- To attain this goal, using graph-learning based approaches can be a way forward. We leave this as a future work.

Thank you

Thank you

Let us know if you have any questions. Also feel free to contact at anindyam.jan@gmail.com if you are interested in our work and want to explore more!