Sparsity-based Online Missing Sensor Data Recovery

Di Guo
Xiamen University, China

Xiaobo Qu, Lianfen Huang, Yan Yao, Zicheng Liu, Ming-Ting Sun

May 22, 2012
Contents

- Missing data recovery
- Sparsity-based recovery model
- Dictionary design
- Two extensions
- Simulation results
- Conclusion & Future work
Wireless Sensor Networks

Applications: environment sensing, building, agricultural surveillance, medical care, military
Data is missing

- Node power outage
- Hardware dysfunction
- Channel fading
- Bad environment

dark blue represent missing data

spatial-temporal sampling model
Missing data recovery

- **Retransmission**: not suitable to delay sensitive applications
- **Interpolation methods**: typical ones
  - (1) K-Nearest-Neighbor (KNN)
  - (2) Kriging

Intercommunity: linear combination of available data

Different weight:
- KNN: distance between neighbors;
- Kriging: data statistics (variogram)

Proposed method

- Sparse linear combination of atoms
  \[ x = \Psi \alpha = \sum_{j=1}^{N} \psi_j \alpha_j \]
- Weight relies on the available data
Sparsity

\[ \| \alpha \|_0 \ll N, \; x \in \mathbb{R}^N \]

\[ S/N = 240/4096, \quad \frac{\| \alpha - \alpha_S \|_2^2}{\| \alpha \|_2^2} < 10^{-5} \]
Model

\[ f_n = \Phi_n \alpha_n \]

\[ \begin{pmatrix} f_n^\Lambda_n \\ f_n^\overline{\Lambda}_n \end{pmatrix} = \begin{pmatrix} \Phi_n^\Lambda_n \\ \Phi_n^\overline{\Lambda}_n \end{pmatrix} \alpha_n \]

\[ \arg \min_{\alpha_n} \|\alpha_n\|_1 \quad \text{s.t.} \quad f_n^\Lambda_n = \Phi_n^\Lambda_n \alpha_n \]

Assumption: Gaussian noise

\[ \hat{\alpha}_n = \arg \min_{\alpha_n} \frac{1}{2} \left\| f_n^\Lambda_n - \Phi_n^\Lambda_n \alpha_n \right\|_2^2 + \lambda \left\| \alpha_n \right\|_1 \]

Output: \[ \Lambda_n = \Phi_n \hat{\alpha}_n \]

Maximum a posteriori probability

Key: How to reduce recovery error?
   (1) Dictionary, (2) Available data consistency
Features of WSN data

- smooth, few boundaries
- weak spatial correlation
- strong temporal correlation

Example: surface sunshine duration

- Spatial domain: DCT basis
- Temporal spatial domain: a few past frames + DCT basis (overcomplete dictionary)
Sparsity-based online data recovery

**Proposed approach:**
Sparse Recovery using Overcomplete Dictionary (SROD):
Using a sparse linear combination of the overcomplete dictionary to represent the current frame.

**Motivation:**
temporal correlation among frames
Temporal correlation of frames

Coefficients

Thermal image of a data center
(data from Microsoft)
### Simulation

<table>
<thead>
<tr>
<th>Methods</th>
<th>KNN</th>
<th></th>
<th></th>
<th>SROD</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>20%</td>
<td></td>
<td>10%</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>MAE_frame</td>
<td>1.31</td>
<td>1.40</td>
<td>1.54</td>
<td>1.88</td>
<td>0.88</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(32.8%)</td>
<td>(20.7%)</td>
</tr>
<tr>
<td>MAE_node</td>
<td>1.48</td>
<td>1.48</td>
<td>1.75</td>
<td>1.80</td>
<td>1.06</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(28.4%)</td>
<td>(18.2%)</td>
</tr>
<tr>
<td>RMSE_frame</td>
<td>0.66</td>
<td>0.69</td>
<td>0.66</td>
<td>0.78</td>
<td>0.43</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(34.8%)</td>
<td>(19.7%)</td>
</tr>
<tr>
<td>RMSE_node</td>
<td>0.68</td>
<td>0.67</td>
<td>0.69</td>
<td>0.77</td>
<td>0.43</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(36.8%)</td>
<td>(23.5%)</td>
</tr>
<tr>
<td>Total RMSE</td>
<td>0.76</td>
<td>0.75</td>
<td>0.73</td>
<td>0.84</td>
<td>0.47</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(38.2%)</td>
<td>(25.0%)</td>
</tr>
</tbody>
</table>

**3D-KNN**: anisotropic temporal spatial correlation

**Data missing rate**: 10%, 20%

**Burst missing length**: 5, 10
Robustness to Noise

![Graph showing robustness to noise](image)

- **Total RMSE vs. SNR (dB)**: The graph on the left shows the total root mean square error (RMSE) vs. signal-to-noise ratio (SNR) in decibels (dB). As the SNR increases, the total RMSE decreases, indicating improved robustness to noise.

- **Total RMSE vs. Number of past frames**: The graph on the right demonstrates the total RMSE vs. the number of past frames. Different curves represent different noise levels (Noiseless, 30dB, 20dB, 10dB), with lower RMSE values indicating better performance under varying noise conditions.
Error propagation

- **Problem:** the recovery error of last frame may propagate
- **Possible solution:**
  Leverage the available data of current frame to correct the recovery error in the last frame in some degree.

![Diagram showing error propagation]

Last frame missing, but current frame available
Recovery with Corrected Dictionary (RCD)

Neighboring data consistency

\[
\hat{\alpha}_{n-1} = \arg \min_{\alpha_{n-1}} \frac{1}{2} \left\| f_{n-1}^\Lambda_{n-1} - \Phi_{n-1} \alpha_{n-1} \right\|^2_2 \quad \text{+} \quad \frac{\mu^2}{2\sigma_n^2} \left\| f_n^{\Lambda_{n-1} \cap \Lambda_n} - \Phi_{n-1} \Lambda_n \alpha_{n-1} \right\|^2_2
\]

Update one atom of the dictionary

\[
B_{n-1} = \Phi_{n-1} \hat{\alpha}_{n-1}
\]

Dictionary Update

B_{n-1} \quad B_n \quad B_{n+1} \quad B_{n+2} \quad \cdots

f_{n-1} \quad f_n \quad f_{n+1} \quad f_{n+2} \quad \cdots

A'_{n-1} \quad A'_n \quad A'_{n+1} \quad A'_{n+2} \quad \cdots

Output

B_{n-1}: updated last frame using RCD

A'_n: recovered current frame using SROD
Recovery with future frame compensation (RFFC)

- If delay is not a major concern:
  
  \[
  \hat{\alpha}_n = \arg \min_{\alpha_n} \frac{1}{2} \left\| f_n - \Phi_n \alpha_n \right\|_2^2 + \lambda \left\| \alpha_n \right\|_1 + \frac{\mu}{2\sigma_n^2} \left\| \frac{f_{n+1}}{\sigma_{n+1}} - \Phi_n \alpha_n \right\|_2^2
  \]

  Neighboring data consistency

  Current frame \( B_n = \Phi_n \hat{\alpha}_n \)

  Dictionary Update

  \( B_{n-1} \) \( B_{n-2} \) \( \vdots \) \( B_{n-L} \)

  \( f_{n-1} \) \( f_n \) \( f_{n+1} \) \( f_{n+2} \) \( \cdots \)

  Output \( B_n \): recovered current frame

  delay

  \( B_{n+1} \) \( B_{n+2} \) \( B_{n+3} \) \( \cdots \)
Three proposed sparsity-based recovery method compare with corresponding 3-D KNN

Missing rate: 20%, burst missing length: 1

<table>
<thead>
<tr>
<th>Methods</th>
<th>KNN</th>
<th>KNN-CD</th>
<th>KNN-FFC-1</th>
<th>SROD</th>
<th>RCD</th>
<th>RFFC-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>KNN</td>
<td>KNN-CD</td>
<td>KNN-FFC-1</td>
<td>SROD</td>
<td>RCD</td>
<td>RFFC-1</td>
</tr>
<tr>
<td>MAE_frame</td>
<td>1.55</td>
<td>1.54 (0.6%)</td>
<td>1.46 (5.8%)</td>
<td>0.97 (37.4%)</td>
<td>0.95 (38.7%)</td>
<td>0.79 (49.0%)</td>
</tr>
<tr>
<td>MAE_node</td>
<td>1.80</td>
<td>1.79 (0.6%)</td>
<td>1.73 (3.9%)</td>
<td>1.25 (30.6%)</td>
<td>1.24 (31.1%)</td>
<td>1.08 (40.0%)</td>
</tr>
<tr>
<td>RMSE_frame</td>
<td>0.66</td>
<td>0.66 (-)</td>
<td>0.62 (6.1%)</td>
<td>0.38 (42.4%)</td>
<td>0.37 (43.9%)</td>
<td>0.30 (54.5%)</td>
</tr>
<tr>
<td>RMSE_node</td>
<td>0.69</td>
<td>0.69 (-)</td>
<td>0.65 (5.8%)</td>
<td>0.39 (43.5%)</td>
<td>0.38 (44.9%)</td>
<td>0.32 (53.6%)</td>
</tr>
<tr>
<td>Total RMSE</td>
<td>0.72</td>
<td>0.72 (-)</td>
<td>0.69 (4.2%)</td>
<td>0.43 (40.3%)</td>
<td>0.42 (41.7%)</td>
<td>0.36 (50.0%)</td>
</tr>
</tbody>
</table>

- Error reduce by 40%
- RFFC reduce error by 10% over SROD
Conclusion

❖ Propose sparsity-based online data recovery method
❖ Construct an overcomplete dictionary: past frames + DCT basis
❖ Recovery performance significantly outperforms KNN
❖ Robust to certain noise
❖ RCD may reduce error propagation
❖ RFFC can further improve recovery performance
Future work

- Test missing pattern from the perspective of wireless communication
- Extract data feature using data mining
- Design dictionary and optimization algorithms
Acknowledgement

Data from
Dr. Jie Liu in Microsoft

Funds from
• Tsinghua-Qualcomm Joint Research Program
• National Natural Science Foundation of China (No. 61001142)
• China Scholarship Council
Thank you

Any questions?