
Zitao Liu
University of Pittsburgh
ztliu@cs.pitt.edu

November 16, 2015

† This is the joint work with Chris Yan, Jimmy Yang and Milos Hauskrecht.
What is **hierarchical time series** (HTS)? Yahoo! web pages are arranged in certain hierarchy and their **daily page views** become a hierarchical time series.
Hierarchical Time Series

Time series are organized in a hierarchical tree structure and they are consistent between hierarchy levels.

Consistent: Parent = Child 1 + Child 2 + Child 3 + … + Child n
Motivation

Why we care about modeling HTS?

- Resource management.
- User behaviors understanding.
- Advertisement pricing policy.
Problem & Goal

However, missing values occur:

- machine failures
- networking disturbances
- human mistakes

Missing values will contaminate other time series through the hierarchy consequentially.
<table>
<thead>
<tr>
<th></th>
<th>Parent</th>
<th>Child 1</th>
<th>Child 2</th>
<th>Child 3</th>
<th>Child 4</th>
<th>Child 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>$t_1$</td>
<td>$t_2$</td>
<td>$t_3$</td>
<td>$t_4$</td>
<td>$\cdots$</td>
<td></td>
</tr>
</tbody>
</table>

Accurately estimate the missing values.

\[ s.t \]

Estimation is hierarchically consistent.
In this work, we develop a new missing value estimation algorithm \textit{HTSIImpute} which

- utilizes the temporal dependence information within each individual time series (\textit{LOcal regrESSion (LOESS)})

- exploits the intra-relations between different time series (\textit{Subspace Projection})

- guarantees hierarchical consistency (\textit{Hierarchical Consistency Projection})
Use LOESS to initially estimate the missing values.

Advantages:
- nonparametric
- robust
- locally weighted
HTSImpute - Subspace Projection

Full rank matrix $\xrightarrow{P_{r-SVD}(\cdot)}$ Low rank matrix

Rank: $n$ $\xrightarrow{}$ Rank: $l$
\[ Y = \Omega \cdot \hat{L} + \epsilon_Y \]

where \( \hat{L} \) is the “true” estimate of all leaf time series.

We define the hierarchical consistency projection operator using ordinary least square as follows:

\[ P_{HTS}(Y, \Omega) = \Omega \hat{L} = \Omega (\Omega^\top \Omega)^{-1} \Omega^\top Y \]
HTSImpute - Idea Illustration

Start

LOESS Initialization

Goal
HTSImpute - Idea Illustration

Start

HTS

Subspace Projection

Hierarchical Consistent Projection

Goal

LOESS Initialization

Subspace

HTS
HTSImpute - Idea Illustration

Goal

Start

HTS

HTS

Subspace

Subspace Projection

Hierarchical Consistent Projection

LOESS Initialization

Subspace
Figure 1: Synthetic data.

Figure 2: Yahoo! web traffic data.
**Avg-MAPE**: measures the estimation accuracy.

\[
\frac{\text{Estimated Value} - \text{True Value}}{\text{True Value}}
\]

**Avg-HCG**: measures the hierarchical consistency.

“Estimated Parent Value – Sum of Estimated Child Values”
Experiments - Baseline

Regression Methods
- Local regression (LOESS)

Subspace Methods
- Matrix Factorization (MF)
- Matrix Completion (MC) using softImpute
- Weighted Low Rank Approximation (wLRA)

Latent Variable Models
- Probabilistic PCA (pPCA)
### Experiments - Results

**Table 1:** Avg-MAPE results on *FP* dataset.

<table>
<thead>
<tr>
<th># MP (%)</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOESS</td>
<td>11.57</td>
<td>11.84</td>
<td>11.66</td>
<td>11.78</td>
<td>11.82</td>
<td>11.78</td>
</tr>
<tr>
<td>NMF_Euclidean</td>
<td>21.19</td>
<td>17.75</td>
<td>16.49</td>
<td>17.41</td>
<td>15.70</td>
<td>15.97</td>
</tr>
<tr>
<td>MC</td>
<td>57.30</td>
<td>44.27</td>
<td>46.76</td>
<td>57.11</td>
<td>63.37</td>
<td>66.85</td>
</tr>
<tr>
<td>pPCA</td>
<td>101.51</td>
<td>100.17</td>
<td>100.21</td>
<td>100.03</td>
<td>100.01</td>
<td>100.19</td>
</tr>
<tr>
<td>wLRA</td>
<td>42.39</td>
<td>73.64</td>
<td>96.91</td>
<td>41.54</td>
<td>34.03</td>
<td>30.10</td>
</tr>
<tr>
<td>HTSImpute</td>
<td><strong>7.19</strong></td>
<td><strong>7.48</strong></td>
<td><strong>7.40</strong></td>
<td><strong>7.87</strong></td>
<td><strong>8.22</strong></td>
<td><strong>8.36</strong></td>
</tr>
</tbody>
</table>
### Table 2: Avg-HCG results on FP dataset (log$_{10}$ scale).

<table>
<thead>
<tr>
<th># MP (%)</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOESS</td>
<td>-2.61</td>
<td>-2.14</td>
<td>-1.93</td>
<td>-1.67</td>
<td>-1.54</td>
<td>-1.45</td>
</tr>
<tr>
<td>NMF_KL</td>
<td>-2.52</td>
<td>-2.04</td>
<td>-1.95</td>
<td>-1.71</td>
<td>-1.60</td>
<td>-1.51</td>
</tr>
<tr>
<td>NMF_Euclidean</td>
<td>-2.01</td>
<td>-1.88</td>
<td>-1.80</td>
<td>-1.62</td>
<td>-1.55</td>
<td>-1.47</td>
</tr>
<tr>
<td>MC</td>
<td>-1.91</td>
<td>-1.35</td>
<td>-0.96</td>
<td>-0.19</td>
<td>0.30</td>
<td>1.34</td>
</tr>
<tr>
<td>pPCA</td>
<td>0.79</td>
<td>0.97</td>
<td>1.39</td>
<td>1.49</td>
<td>2.18</td>
<td>1.78</td>
</tr>
<tr>
<td>wLRA</td>
<td>-5.65</td>
<td>-4.83</td>
<td>-1.58</td>
<td>-1.67</td>
<td>-1.55</td>
<td>-1.47</td>
</tr>
<tr>
<td>HTSIImpute</td>
<td><strong>-16.66</strong></td>
<td><strong>-16.32</strong></td>
<td><strong>-16.04</strong></td>
<td><strong>-15.80</strong></td>
<td><strong>-15.71</strong></td>
<td><strong>-15.61</strong></td>
</tr>
</tbody>
</table>
Conclusion

In this work, we have presented an algorithm for HTS missing value estimation, specializing in:

- taking advantage of temporal dependence information within each individual time series.
- utilizing intra-relations between different time series across the hierarchy.
- providing high satisfaction of the hierarchical consistency.
Thank you

November 16, 2015
HTSIImpute - Hierarchical Consistency Projection

\[
\begin{bmatrix}
1 & 1 \\
1 & 0 \\
0 & 1
\end{bmatrix}
\]

Summing matrix \( \Omega \)
HTSImpute - Hierarchical Consistency Projection

\[ \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} + \text{Noise} \]

Summing matrix $\Omega$