Preprocessing Time Series Data with MATLAB

This reference shows common use cases, but is by no means comprehensive. The icon provides links to relevant sections of the MATLAB documentation to learn more.

**Timetable**
MATLAB datatype designed to organize and work with time series data.

**Components of a Timetable**
- **Times of Day/Times**
- **Variable Names**
- **Data**
  - `double`, `int`, logical, char, string, cell, and other arrays
- **duration** (offset from a reference time)

Create Timetables

\[
\text{tt} = \text{table2timetable}(t);
\]
(First datetime or duration variable in "t" becomes the row times.)

**Timetable Manipulation**

**Access Data**
These return the same array:
- \( \text{tt}.\text{Temperature} \)
- \( \text{tt}(:,\text{‘Temperature’}) \)
- \( \text{tt}(:,1) \)

Add a New Variable

\[
\text{tt}.\text{newVar} = \text{zeros}(\text{height}('t'),1);
\]

**Data Cleaning**

**Smooth Data**

\[
B = \text{smoothdata}(A,\text{method});
\]

Smooth noisy data with methods:
- 'movmean', 'movmedian', 'gaussian', 'lowess', 'loess', 'rloess', 'sgolay' or

**Detect Outliers**

\[
\text{TF} = \text{isoutlier}(A,\text{method});
\]

Identify outliers with methods:
- 'median', 'mean', 'quartiles', 'grubbs', 'gesd'

**Detect Change Points**

\[
\text{TF} = \text{ischange}(A,\text{method});
\]

Find abrupt changes with methods:
- 'mean', 'variance', 'linear'

**Merge Timetables**
Synchronize multiple timetables to a common time vector.

\[
\text{tt} = \text{synchronize}(\text{tt1},\text{tt2},...,\text{ttN});
\]

Synchronizing often results in missing data points (times at which a variable was not measured). `synchronize` supports several methods for adjusting data to fill in gaps:

- Fill: 'fillwithmissing', 'fillwithconstant'
- Interpolation: 'linear', 'spline', 'pchip'
- Nearest Neighbor: 'previous', 'next', 'nearest'
- Aggregation: 'mean', 'min', 'max', @func...

**Missing Data**

Find Missing Values

\[
\text{TF} = \text{ismissing}('t');
\]

Fill Missing Values

\[
\text{tt} = \text{fillmissing}(\text{tt},\text{method});
\]
Replace missing values with values calculated from nearby points with methods:
- 'previous', 'next', 'nearest', 'linear', 'spline', 'pchip'

Remove Rows Containing Missing Values

\[
\text{tt} = \text{rmmissing}('t');
\]

**Big Data**
Tall arrays extend MATLAB functions to work on data too big to load into memory.

Create a "tall" timetable:

\[
\% \text{Create a datastore that points to the data}
\text{ds} = \text{datastore('*.csv');}
\%
\% \text{Create a tall table from the datastore}
\text{t} = \text{tall}(\text{ds});
\%
\% \text{Convert to a timetable}
\text{tt} = \text{table2timetable}(\text{t});
\]

© 2018 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.