Time is on your side: Semantic enrichment of time series

Bojan Božić

Dublin Institute of Technology

What is a time series?

- What we need: value(s) + time stamp
- What would be grand: constant time interval
- Mathematically: values $Y_1, Y_2, ...$ of a variable Y at points in time $t_1, t_2, ...$ Therefore: $Y = F(t)$

Example

Air quality sensor measuring carbon monoxide and ozone concentrations in air (in ppm), and sending in values every 10 minutes.

Subdisciplines and Goals

We distinguish:

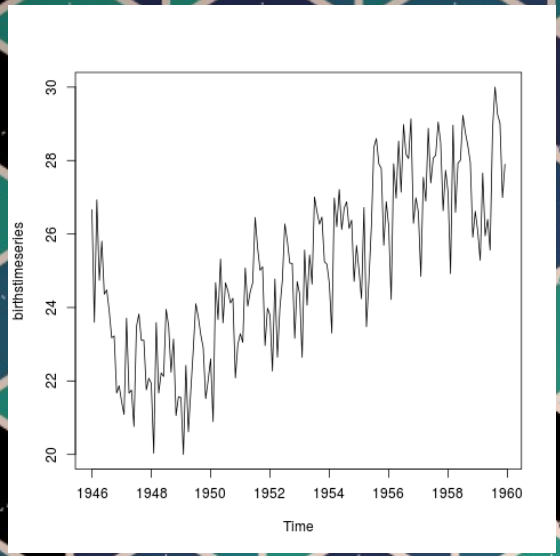
- Time Series Analysis: Analysing time series data in order to extract meaningful statistics and other characteristics of the data.
- Time Series Forecasting: Estimating many future aspects of a business or other operation based on the current time series.

Which aim at:

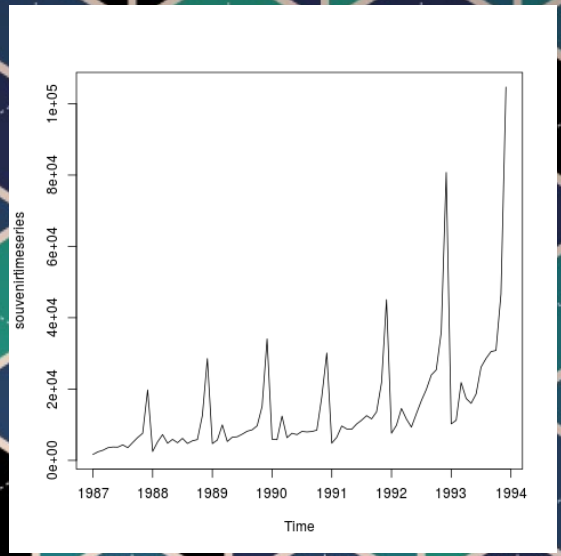
- Identifying the nature of the phenomenon represented by the sequences of operations.
- Predicting future values of the time series variables.

Patterns

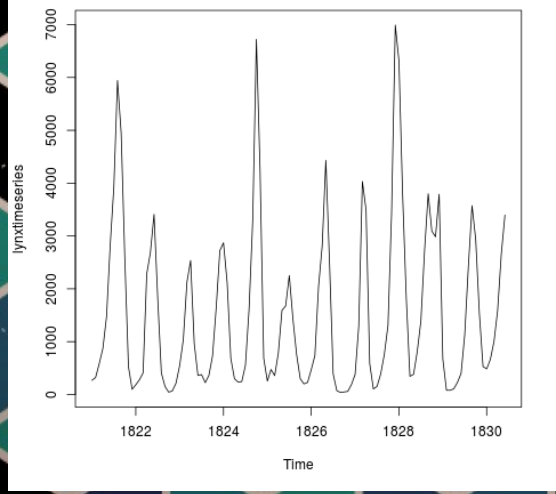
Trend:



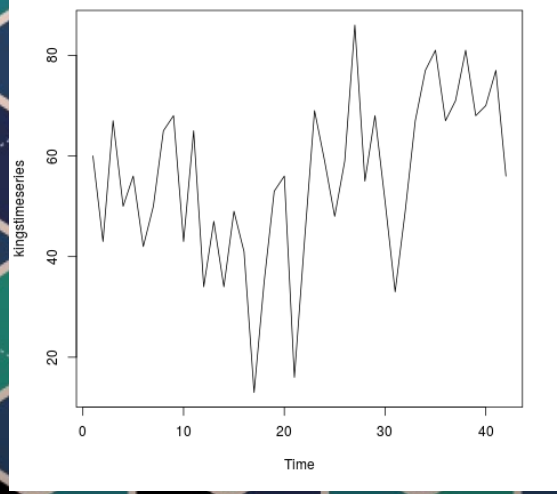
Seasonal:



Cyclical:



Irregular:



Decomposition

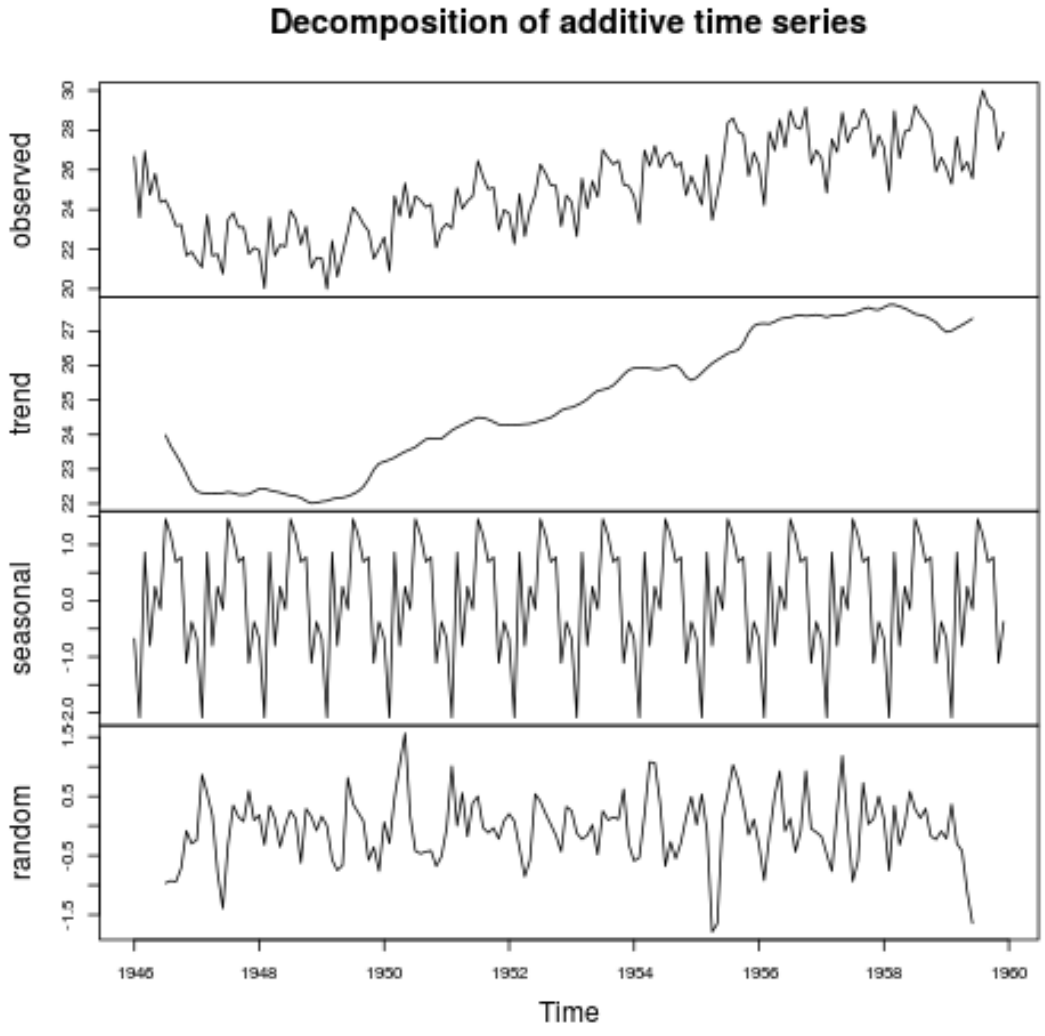


Figure 1: Decomposed births time series.

Mathematical Models

Additive Model:

1. Assume data is sum of components: $Y_t = T + S + C + I$
2. If one of the components is not contained the value is zero, therefore: $Y_t = T + S + I$
3. Seasonal component is independent of trend $\Longrightarrow$ magnitude of seasonal swing is constant

Multiplicative Model:

1. Assume data is product of components: $Y_t = T * S * C * I$
2. If one of the components is missing, the value is assumed to be 1: $Y_t = T * S * I$
3. Seasonal factor is a ratio to the trend $\Longrightarrow$ magnitude increases or decreases according to trend

Smoothing

- Removes random variation and shows trends and cycles
- Two methods:
  1. Averaging Smoothing: e.g. $SMA_i = \frac{\sum_{k=i-n}^{i} x_k}{n}$
  2. Exponential Smoothing: e.g.
     $EMA_i = EMA_{i-1} + \alpha * (x_i - EMA_{i-1})$
     $\alpha = \frac{2}{n+1}$

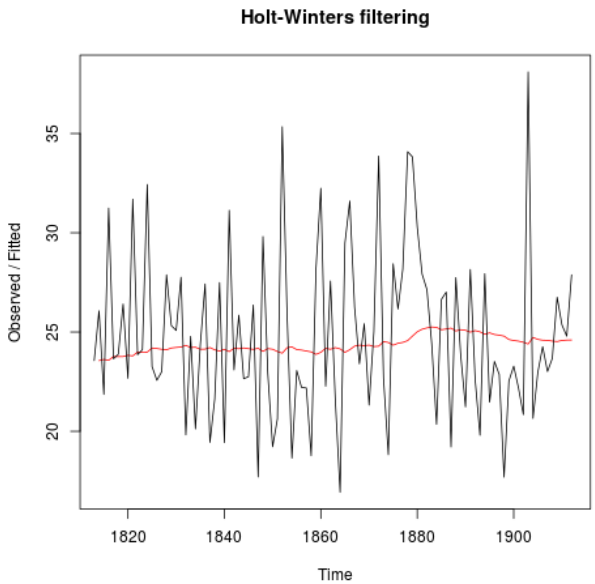


Figure 2: Smoothed forecasts for rainfall using Holt-Winters (Simple Exponential Smoothing)

Why add Semantics?

- Experts in different fields have to deal with time series (environment, finance, etc.)
- Time series represent measurements with a lot of data (values for years with measurements for every second)
- A lot of time and expert knowledge required to find the right part of a time series
- Time series processors are highly customized and developed exclusively for one purpose

Expressions

Expression	**Meaning**
`< [n].sin * 2 + 3 >`	Calculation is applied to all slots.
`A, B < A[n] + 2 * B[n] >`	Combination of two time series (aggregation).
`< [n] > every 2 hours`	Projection to a fixed time grid.
`< (t .. t-2).mean > every 1 hour`	Sliding mean value.
`< [n]->hot if [n].temperature > 100 otherwise [n]->cold >`	Filtering, classification.

Table 1: An overview of most commonly used time series processing expressions.

Semantic Time Series

- Enrichment of time series with meta-information (annotations)
- Reduction of processor and language complexity
- Usage of ontologies to define the domain of interest
- Usage of reasoning to generate new meta-information

Comparison

Datatype	Example
Temperature	30.5, 15.2, ...
Unit	°C, °F, ...
Legal Regulations	The threshold of temperature for heat warnings is … Precipitation limit for flooding warnings is ...
Meaning	°C signalizes temperature, L/m² is the unit for precipitation, Ozone is a hazardous material...

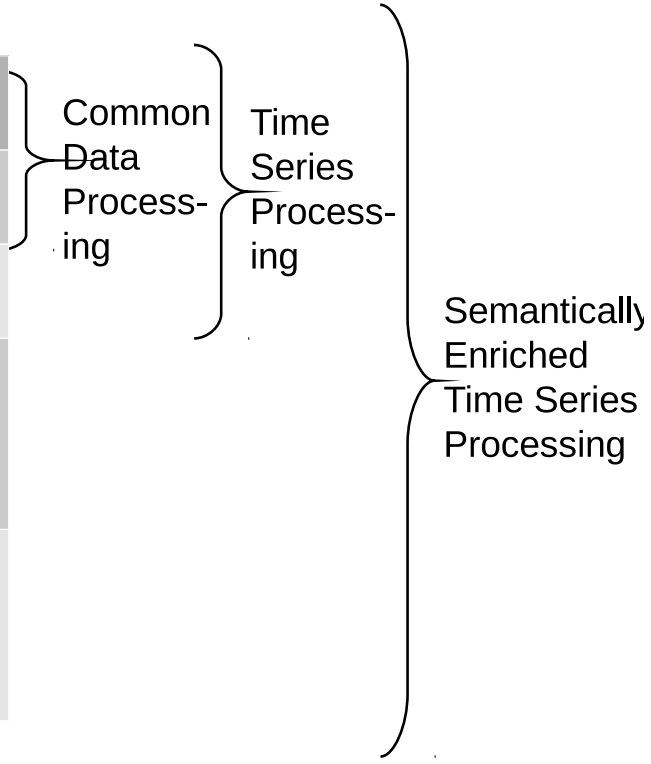


Figure 3: Difference between one-dimensional data, time series, and semantic time series

Sample Semantic Expression

Sample regular Time Series Processing expression:

`MeteoTS < warning if precipitation > 1000 l/m`^{2} `or temperature > 40`$^\circ$`C or wind > 56 knots ... >`

Semantic Time Series Processing expression:

`MeteoTS < warning if value > allowed >`

Ontologies

1. Definition of all common classes and properties, which are valid for all possible time series ontologies.
2. Extraction of a bridge ontology which can be used as a common interface for all domain ontologies.
3. Every individual domain ontology needs to inherit from the bridge ontology (it has to define all classes which are also defined in the bridge ontology).
4. Therefore, an ontology graph can be constructed, which has the bridge ontology in the center.

Time Series Ontology

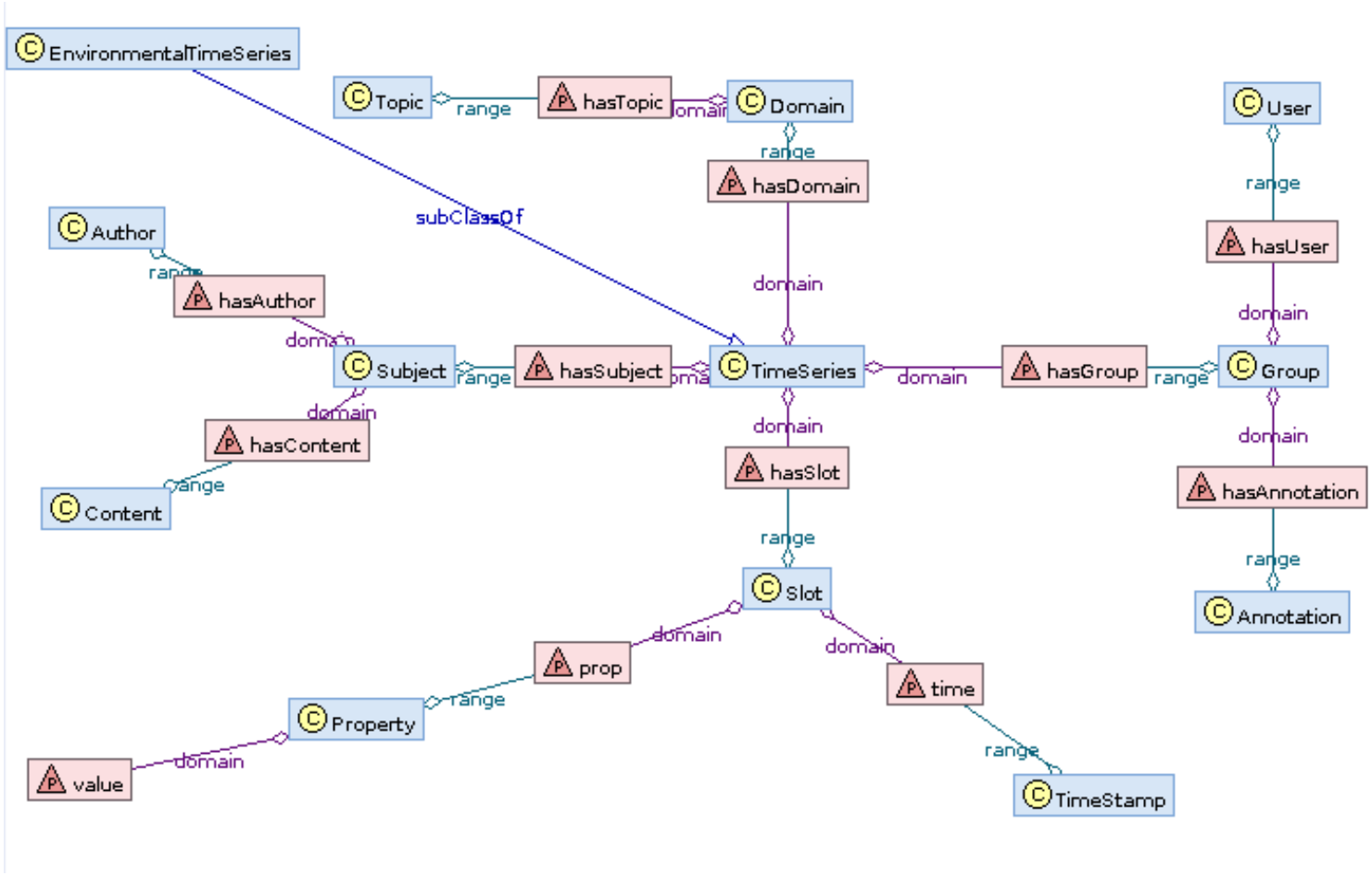


Figure 4: The Time Series Bridge Ontology for connecting domain knowledge.

Contact

Bojan Božić
CeADAR Research Fellow
Email: bojan.bozic@dit.ie
Twitter: @bojan_bozic

TimeSeries data: robjhyndman.com/tsdldata/
Images: suwalls.com, patterncooler.com