H2O Time Series

Using machine learning for solving time series problems

Machine learning with time series

Typical steps:

1. Conversion data to regular time series (if needed)
    - Sampling rate or period to use
    - Convert irregular time data (or events) into regular sampled
    - Resample (in time) and convert to numeric categorical/symbolic variables
2. Dynamic preprocessing (optional): convert dynamic data into static pattern
    - Tapped delay line (time shift), fixed
    - Apply dynamic filter (time convolution), adjustable
    - Use transformed domain
3. Apply machine learning
    - Static (standard) if dynamic processing is available
    - Dynamic models: recurrent neural networks and variants

Machine learning with time series

Conversion data to regular time series

- Define convenient sampling rate or period (time resolution), according to problem need and variability (analyze autocorrelation)
- Regularize time series:
  - Convert irregular time data (or events) into regular sampled
  - Complete when needed
  - Convert categorical/symbolic into numeric (optional)
- Reshape :
  - Under sampling: aggregation
  - Over sampling: smoothing

Machine learning with time series

Dynamic preprocessing

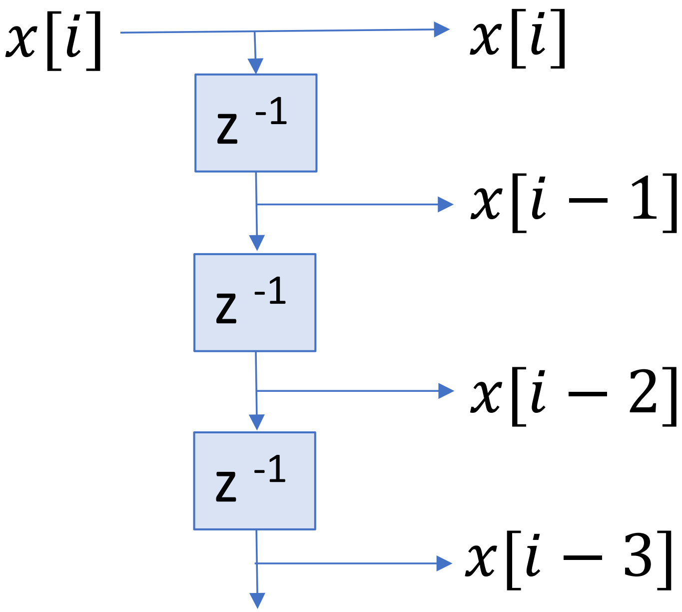
- Needed for static machine learning model to deal with past information
- Simple processing may be convenient: smoothing, differentiating, etc.
- Other transformation may be used (non linear like log, etc.)
- Complex dynamic processing (array):
  - Time domain:
    - Tapped delay line: time shifts (past information over a window)
    - Dynamic filter array: time convolution (may be adjustable)
  - Transformed domain (over window): FFT, wavelet transform, etc.
  - Result: each time series variable converted in array suitable for input at the machine learning process

Machine learning with time series

Dynamic processing array

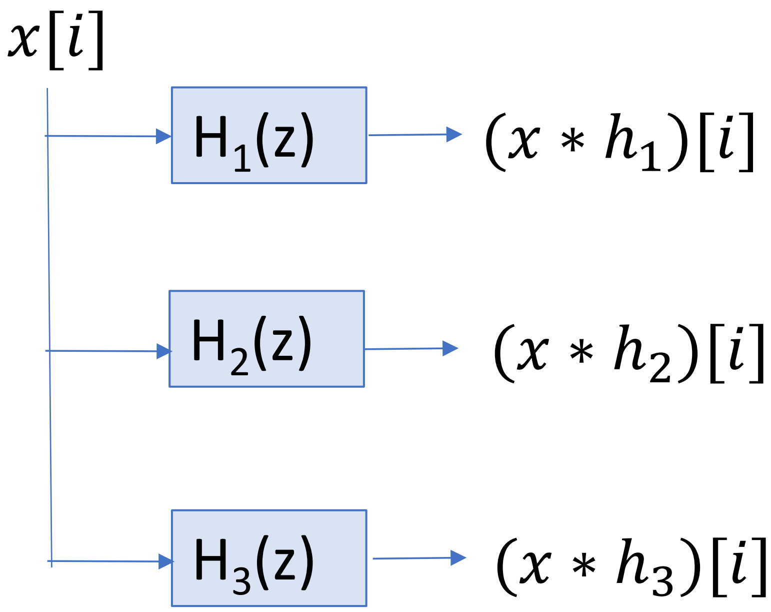
Tapped delay line

Time shift



Dynamic filters array

Time convolution



Machine learning with time series

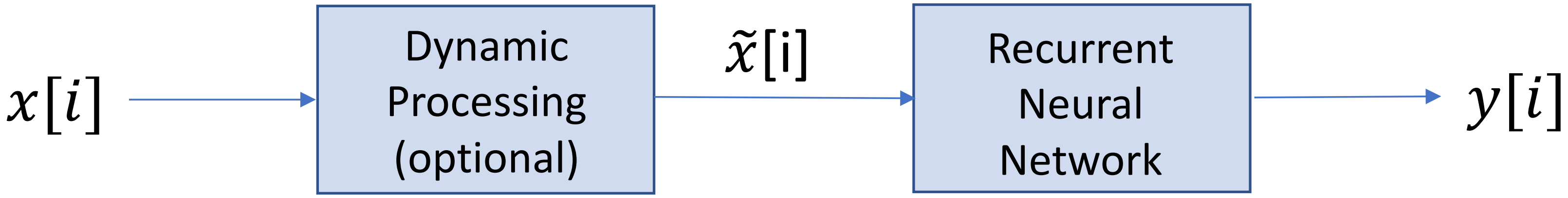
Application of machine learning

- Style of learning:
  - *Long term* for fixed model: single process off-line training, then use with no adjustment
  - *Short term* for (time) variant model: initial off-line training (optional) followed by use with on-line training (adaptive)
- Type of machine learning model:
  - *Static*: output is function of present input only, dynamic aspect of times series must be captured by dynamic preprocessing. Most machine learning models.
  - *Dynamic*: output is function of actual input and past history. Recurrent neural networks (RNN) and variants (Long short-term memory neural network, etc.)

Machine learning with time series

Machine Learning with dynamic preprocessing

Machine Learning with dynamic system (like recurrent neural network)



Machine learning with time series

Use cases

- Prediction or forecasting:
  - Given a time series (and optional inputs), produce expected future value(s):
    $$\{x[-\infty{:}\, i], v[-\infty{:}\, i]\} \Longrightarrow \tilde{x}[i+d],\ \ d \geq 1$$
  - Examples:
    - Simple one-step forecasting (single variable)
      $$x[-\infty{:}\, i] \ \Longrightarrow \tilde{x}[i+1]$$
    - Soft sensor: target values are known only at training
      $$v[-\infty{:}\, i] \Longrightarrow \ \tilde{x}[i]$$
    - Anomaly detector: detection of unexpected values
      $$\{x[-\infty{:}\, i], v[-\infty{:}\, i]\} \Longrightarrow \varepsilon[i] = x[i] - \tilde{x}[i]$$

Machine learning with time series

Use cases

- Classification:
  - Given a time series with a known classification (supervised)

    $\{x[-\infty: i], c[-\infty: i-1]\} \Longrightarrow \tilde{c}[i]$, training phase

    $x[-\infty: i] \Longrightarrow \tilde{c}[i]$, application phase
  - Given a time series with unknown classification (unsupervised)

    $x[-\infty: i] \ \Longrightarrow \ c[i]$

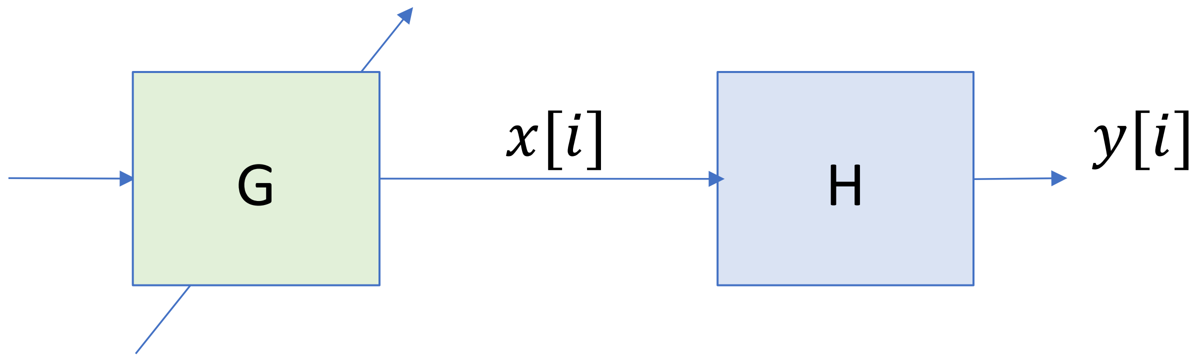
Machine learning with time series

Use cases

- Control or Decision Making
  - For a dynamic system H with input x, output y:

$$x[-\infty : i] \overset{H}{\Rightarrow} y[i]$$

  - Define a controller G to produce suitable input into H:

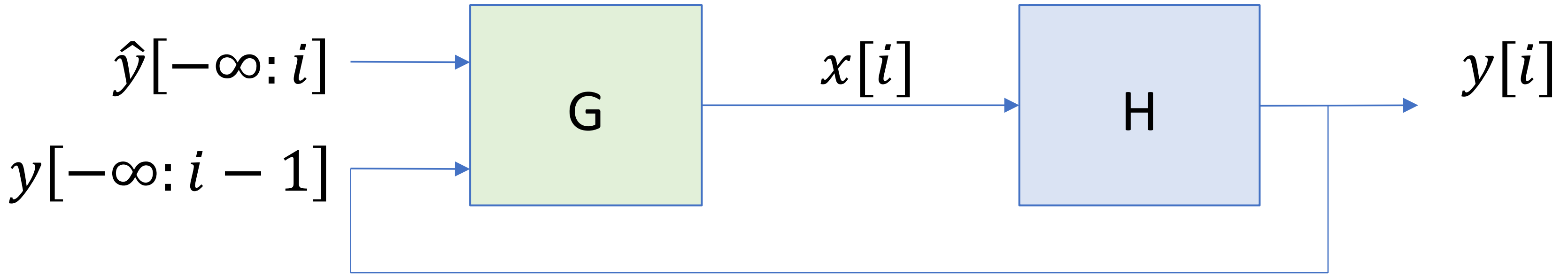


Machine learning with time series

Control or Decision Making

1. *Reference problem*

   Achieve or follow a desired output (target or reference): $\hat{y}$



- Define synthetic model G such that $GoH(\hat{y}) \cong y$:

$$\{\hat{y}[-\infty: i], y[-\infty: i-1]\} \overset{G}{\Rightarrow} x[-\infty: i] \overset{H}{\Rightarrow} y[i] \quad | \; y[i+d] \cong \hat{y}[i+d],\ d \geq 1$$

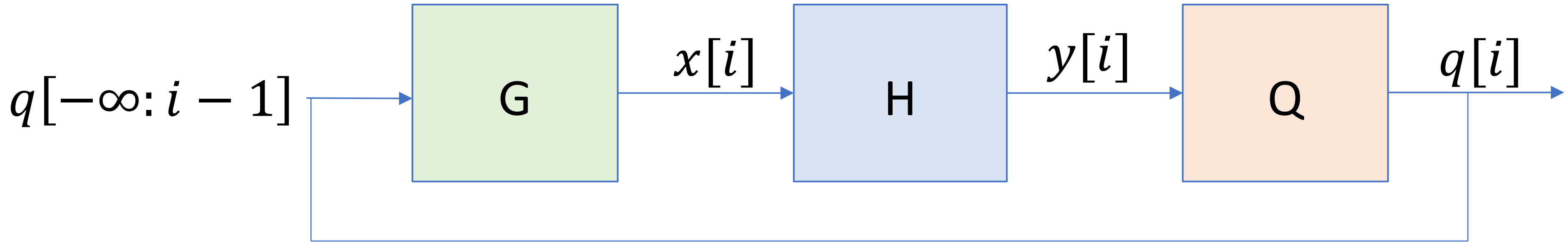
- Train G, may use other model to predict H (predictive control)

Machine learning with time series

Control or Decision Making

2. Optimization problem

Optimize (min or max) an objective function the output: $q = Q(y)$



- Define synthetic model G such that $G \circ H \circ Q(d) = q[i+d]$ be optimal

  $\{q[-\infty: i-1]\} \overset{G}{\Rightarrow} x[-\infty: i] \overset{H}{\Rightarrow} y[-\infty: i] \overset{Q}{\Rightarrow} q[i] \mid \quad Max_G q[i+d]$ (or Min)
- Train G, may use other model to predict H and Q