# Personal Proceeding on Time Series (3)

## --PLSTM, STL and E2E Seasonal Decomposing

## (Mar 22, 2017)

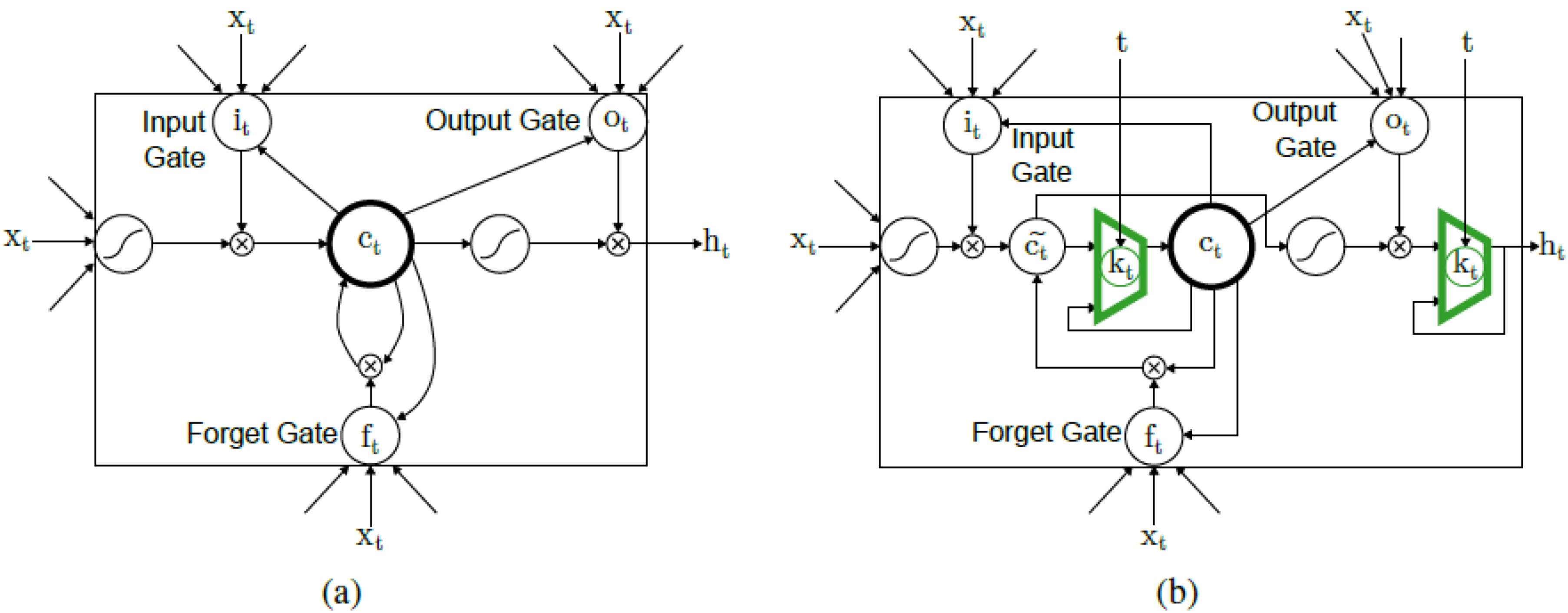
YANG Jiancheng

# Outline

- I. Phased LSTM
- II. Seasonal-Trend Decomposition Based on Loess (STL)
- III. End-to-end Seasonal Decomposing

## • I. Phased LSTM

- **Intuition**

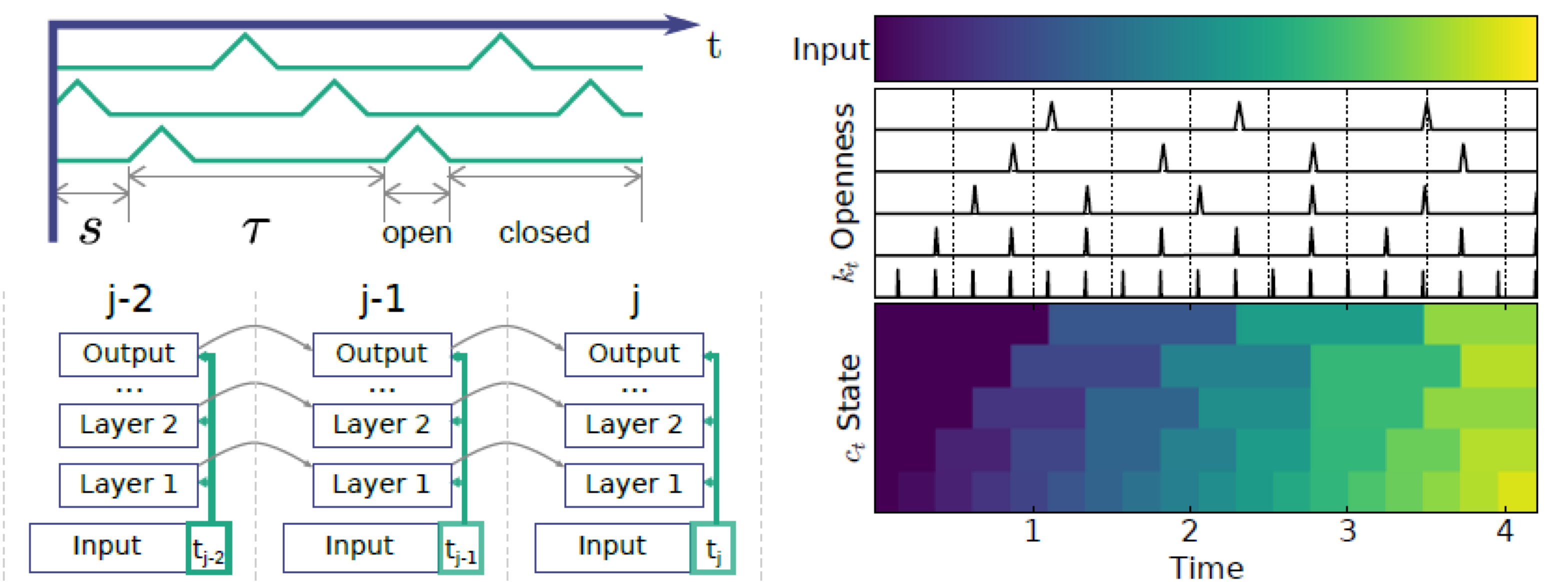



Model architecture. **(a)** Standard LSTM model. **(b)** Phased LSTM model

# • I. Phased LSTM

- **Timing Gate $k_t$**

$$\phi_t = \frac{(t-s) \bmod \tau}{\tau}, \qquad k_t = \begin{cases} \frac{2\phi_t}{r_{on}}, & \text{if } \phi_t < \frac{1}{2}r_{on} \\ 2 - \frac{2\phi_t}{r_{on}}, & \text{if } \frac{1}{2}r_{on} < \phi_t < r_{on} \\ \alpha\phi_t, & \text{otherwise} \end{cases}$$




(a) (b)

# • I. Phased LSTM

- **Cell and Hidden State Update**

Regular LSTM

$$i_t = \sigma_i(x_t W_{xi} + h_{t-1} W_{hi} + w_{ci} \odot c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma_f(x_t W_{xf} + h_{t-1} W_{hf} + w_{cf} \odot c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_c(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \quad (3)$$

$$o_t = \sigma_o(x_t W_{xo} + h_{t-1} W_{ho} + w_{co} \odot c_t + b_o) \quad (4)$$

$$h_t = o_t \odot \sigma_h(c_t) \quad (5)$$

**(3)=>(7,8), (5)=>(9,10)**

PLSTM

$$\widetilde{c_j} = f_j \odot c_{j-1} + i_j \odot \sigma_c(x_j W_{xc} + h_{j-1} W_{hc} + b_c) \quad (7)$$

$$c_j = k_j \odot \widetilde{c_j} + (1 - k_j) \odot c_{j-1} \quad (8)$$

$$\widetilde{h_j} = o_j \odot \sigma_h(\widetilde{c_j}) \quad (9)$$

$$h_j = k_j \odot \widetilde{h_j} + (1 - k_j) \odot h_{j-1} \quad (10)$$

# • I. Phased LSTM

- **Highlights on Memory Keeping**

a) For a regular LSTM, the memory cell will decay due to “forget” exponentially

$$c_n = f_n \odot c_{n-1} = (1-\epsilon) \odot (f_{n-1} \odot c_{n-2}) = \ldots = (1-\epsilon)^n \odot c_0 \quad . \qquad (11)$$

b) While PLSTM can keep well the memory due to the gate closing

c) An example

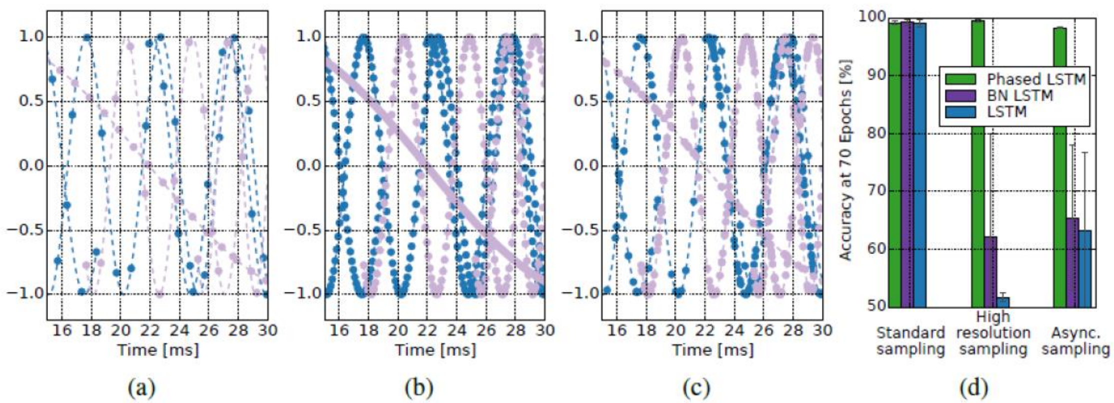



Figure 3: Frequency discrimination task. The network is trained to discriminate waves of different frequency sets (shown in blue and gray); every circle is an input point. **(a)** Standard condition: the data is regularly sampled every 1 ms. **(b)** High resolution sampling condition: new input points are gathered every 0.1ms. **(c)** Asynchronous sampling condition: new input points are presented at intervals of 0.02 ms to 10 ms. **(d)** The accuracy of Phased LSTM under the three sampling conditions is maintained, but the accuracy of the BN-LSTM and standard LSTM drops significantly in the sampling conditions (b) and (c). Error bars indicate standard deviation over 5 runs.

# • II. Seasonal-Trend Decomposition Based on Loess (STL)

a) Basic form

$$Y_v = T_v + S_v + R_v.$$

b) Lots of "messy" operation

c) Algorithm Block

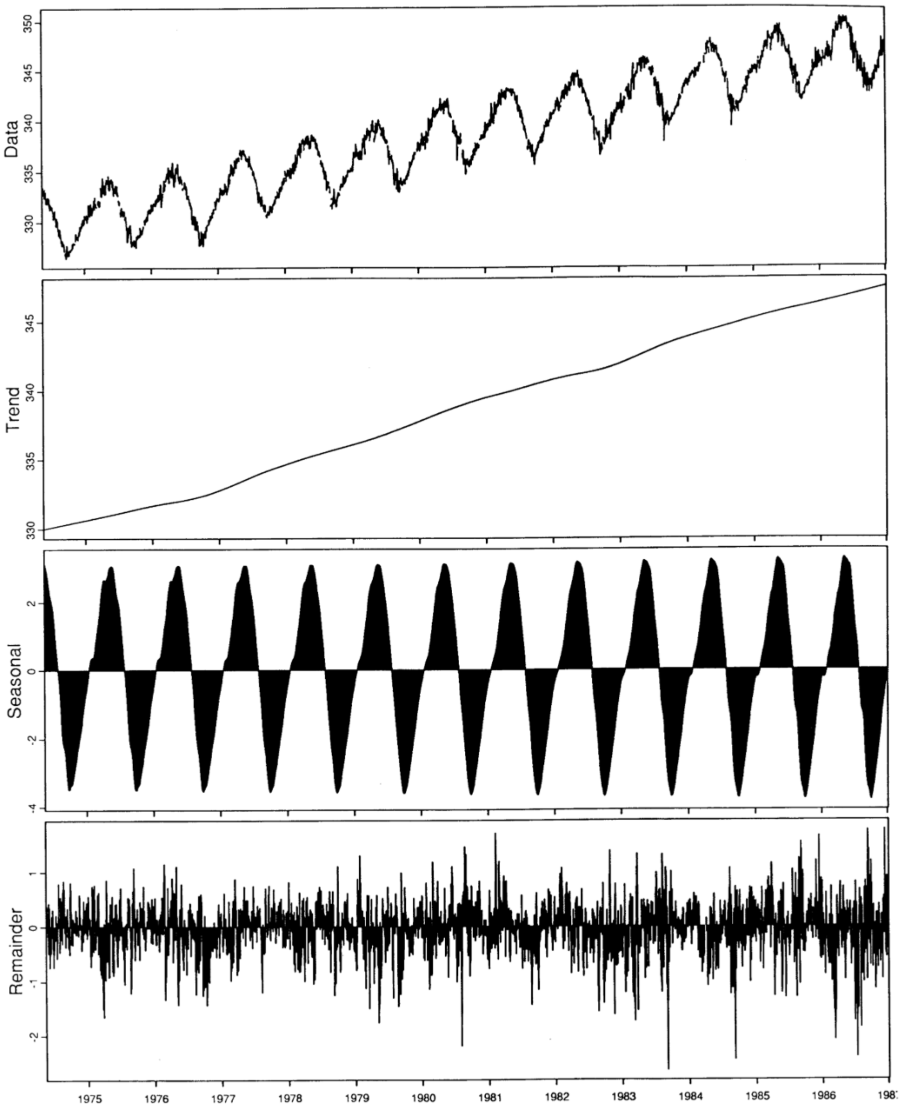



*Fig. 1. Decomposition Plot of Daily Carbon Dioxide Data. The units on the vertical scales are ppm.*

**<u>Algorithm</u>**
Initialization
*for* o=1:outer_step
  *for* I=1:inner_step
    1) detrend
    2) compute seasonal component (loess)
    3) detrend the s using low pass filter
    4) de-seasonality
  compute robust weight
  *end for*
*end for*

- # III. End-to-end Seasonal Decomposing
  - **Not Public**

# Bibliography


- Phased LSTM: Accelerating Recurrent Network Training for Long or Event-based Sequences (NIPS)
- STL: A Seasonal-Trend Decomposition Procedurue Base on Loess (link)

# Thanks for listening!