重慶大學 CHONGQING UNIVERSITY

UNIVERSITY OF Southampton

A Multiple Local Model Learning for Nonlinear and Time-Varying Microwave Heating Process

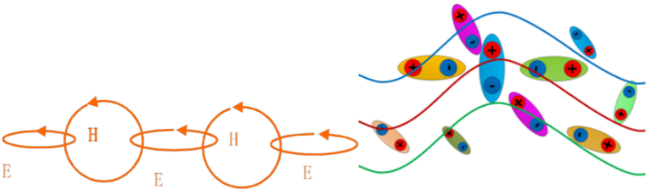
Tong Liu, Shan Liang, Sheng Chen, and Chris J. Harris

tl3n18@soton.ac.uk, lightsun@cqu.edu.cn, sqc@ecs.soton.ac.uk, cjh@chrisharris57@msn.com

International Joint Conference on Neural Networks

July 1, 2019

Microwave heating process



Characteristics

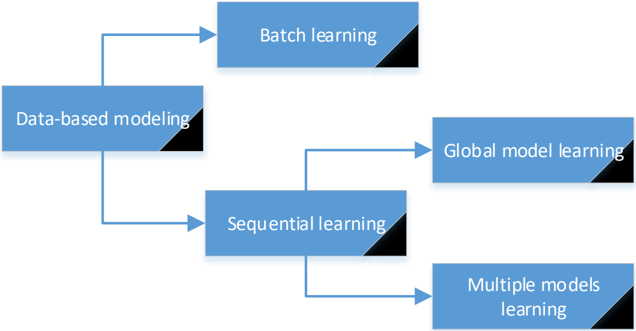
- Generate thermal energy directly due to internal friction of molecules under the effect of electromagnetic field.
- Rapid heat transfer and pollution-free environment.

Obstructions

- Temperature runaway often occurs.
- Temperature model based on first-principle is hard to establish.
- The process is highly nonlinear and time-varying.

Data-based modeling methods

Modeling microwave heating process from data offers a practical alternative.



Batch learning

Model learns from batch of off-line data.

Sequential learning

Model incrementally learns from data streams over time.

Online sequential learning

Online sequential extreme learning machine (OS-ELM) [a]

- Dense hidden nodes are required to cover the input data space.
- Online weight adaptation is computationally costly due to the large model size.
- Only model weights are updated.

[a] A fast and accurate online sequential learning algorithm for feedforward networks, *IEEE Trans. Neural Networks*, Nov. 2006.

Adaptive multiple modelling [a]

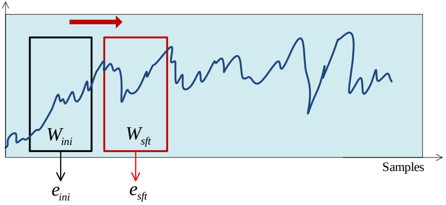
- A set of sub-models are initialised on the same data set.
- The structure of sub-models are fixed during online operation.

[a] A new adaptive multiple modelling approach for non-linear and non-stationary systems, *Int. J. Systems Science*, Oct. 2014.

Problem

Such leaner cannot capture the newly emerged process dynamics.

Adaptation of local linear models



Two null hypotheses

$$H_0^{\mu}: \ \mu_{ini} = \mu_{sft}$$
$$H_0^{\sigma^2}: \ \sigma^2_{sft} = \sigma^2_{ini}$$

- Construct two hypotheses testing to judge whether two data windows are different or not.

Local model construction

$$\widehat{\boldsymbol{y}}_{ini} = f_{ini}(\boldsymbol{X}_{ini}) = \boldsymbol{\Phi\beta}$$
$$\boldsymbol{e}_{ini} = \boldsymbol{y}_{ini} - f_{ini}(\boldsymbol{X}_{ini})$$
$$\boldsymbol{e}_{sft} = \boldsymbol{y}_{sft} - f_{ini}(\boldsymbol{X}_{sft})$$

- Construct local model and estimate errors for consecutive windows.

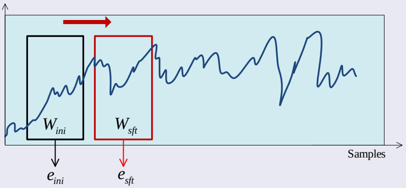
Accepting condition

$$|T| < \lambda_t \text{ and } \chi^2 < \lambda_\chi$$

- W_{ini} and W_{sft} are assumed to be different if this condition is violated, where $T = \sqrt{W}(\mu_{sft} - \mu_{ini})/\sigma_{sft}$, $\chi^2 = (W-1)\sigma^2_{sft}/\sigma^2_{ini}$

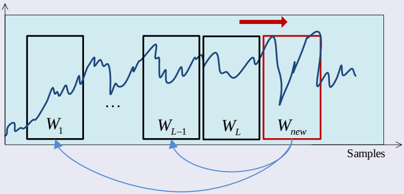
Adaptation of local linear models

New local model detection



- Add new local model if the consecutive windows are significantly different.

Redundant local model deletion



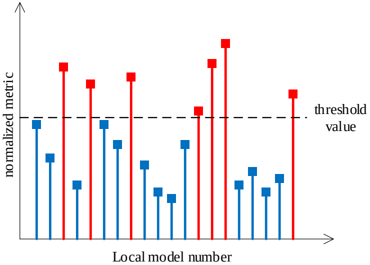
- Delete the local model that is similar to the new model from the local model set $\{f_l\}_{l=1}^{L-1}$.

Advantage

Maintain highly divers and redundancy-free local model set.

Adaptation of model prediction

Selective ensemble learning framework



three questions must be answered

- how to quantify the generalization ability of each local model given the process input $\boldsymbol{x}(t_{next})$?
- how to determine which models should be selected and which models should be filtered out?
- how to make the ensemble of those selected models appropriately?

Adaptation of model prediction

Generalization ability

Modeling error vector of the lth local linear model f_l on the latest p data points $\{\boldsymbol{x}(t-i), y(t-i)\}_{i=0}^{p-1}$

$$e_l(t-i) = y(t-i) - f_l(\boldsymbol{x}(t-i)),\ 0 \leq i \leq p-1$$

Performance metric

$$J_l(t) = \|\boldsymbol{e}_l(t)\|^2$$

Normalize the performance metrics

$$\bar{J}_l(t) = \frac{J_l(t)}{J_{l_{\max}}(t)},\ 1 \leq l \leq L$$

Adaptation of model prediction

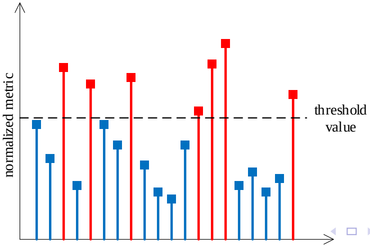
Model selection

The best local model

$$l_{\min} = \arg\min_{1 \le l \le L} \bar{J}_l(t)$$

Other local models whose performance metrics are below a given threshold $0 < \varepsilon \le 1$

$$\Gamma = \{l_1, l_m | 2 \le m \le M, J_{l_m}(t) \le \varepsilon, 2 \le l_m \le L\}$$



Adaptation of model prediction

Ensemble prediction and adaptation

Output estimation based on summation of selected models

$$\widehat{y}(t-i) = \sum_{m=1}^{M} \theta_m(t)\widehat{y}_{l_m}(t-i),\ 0 \leq i \leq p-1$$

$$\text{constraint}: \sum_{m=1}^{M} \theta_m(t) = 1$$

Estimation errors

$$e(t-i) = y(t-i) - \widehat{y}(t-i),\ 0 \leq i \leq p-1$$

Construct the least squares cost function

$$V(t) = \frac{1}{2}\sum_{i=0}^{p-1} e^2(t-i)$$

Adaptation of model prediction

Ensemble prediction and adaptation

Optimization problem

$$\min_{\boldsymbol{\theta}} \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}(t)\bar{\boldsymbol{E}}(t)\boldsymbol{\theta}(t)$$
$$\text{s.t.} \sum_{m=1}^{M} \theta_m(t) = 1$$

Lagrangian function for the optimization

$$L\big(\boldsymbol{\theta}(t);\gamma\big) = \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}(t)\bar{\boldsymbol{E}}(t)\boldsymbol{\theta}(t) + \gamma\big(\mathbf{1}_M^{\mathrm{T}}\boldsymbol{\theta}(t) - 1\big)$$

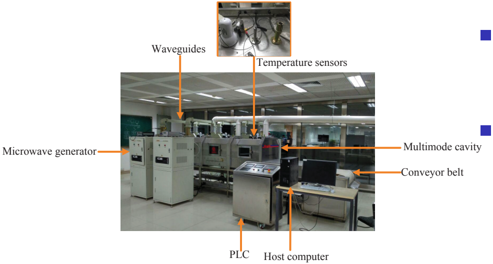
Letting $\frac{\partial}{\partial \boldsymbol{\theta}(t)} L = \mathbf{0}_M$ yields

$$\bar{\boldsymbol{E}}(t)\boldsymbol{\theta}(t) + \gamma\mathbf{1}_M = \mathbf{0}_M$$
$$\widetilde{\boldsymbol{\theta}}(t) = \bar{\boldsymbol{E}}^{-1}(t)\mathbf{1}_M$$

Algorithm summary

- The adaptation of local models can be performed both off-line and on-line.
- The choice of **window size** W trades off the **ability of capturing local characteristics** and the **computational complexity**.
- The choice of **innovation length** p trades off the **computational complexity** and the **robustness against noise**.
- The choice of **threshold** ε trades off the **modeling accuracy** and the **computational complexity**.

Process description



- **System inputs**: microwave powers (five microwave sources), conveyor speed
- **System outputs**: temperature measured by fiber optical sensors at different positions

Mathematical model

$$y(t) = f_{\mathrm{nl-ns}}(\boldsymbol{x}(t); t)$$

$$\text{model inputs}: \boldsymbol{x}(t) = \left[y(t-1)\ \boldsymbol{u}^{\mathrm{T}}(t)\right]^{\mathrm{T}} \in \mathbb{R}^7$$

$$\text{system inputs}: \boldsymbol{u}(t) = \left[u_{p_1}(t)\ u_{p_2}(t)\ u_{p_3}(t)\ u_{p_4}(t)\ u_{p_5}(t)\ v(t)\right]^{\mathrm{T}}$$

Data description

We use datasets from the three sensors, and each data set contains 3,000 data samples. we use the first 1,000 samples for model training, and the last 2,000 samples for online prediction.

Normalization

$$\bar{u}_{p_i}(t) = \frac{u_{p_i}(t)}{1000}, \quad 1 \le i \le 5$$

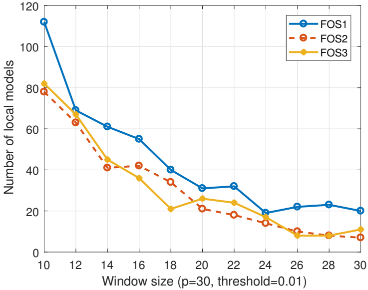
$$\bar{y}(t) = \frac{y(t) - y_{\min}}{y_{\max} - y_{\min}}$$

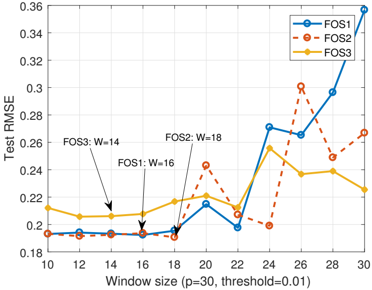
Performance index

$$\mathrm{RMSE}(t) = \sqrt{\frac{1}{t}\sum_{i=1}^{t}\left(y(i) - \widehat{y}(i)\right)^2}$$

$$\mathrm{MAE}(t) = \frac{1}{t}\sum_{i=1}^{t}\left|y(i) - \widehat{y}(i)\right|$$

Parameter sensitive analysis



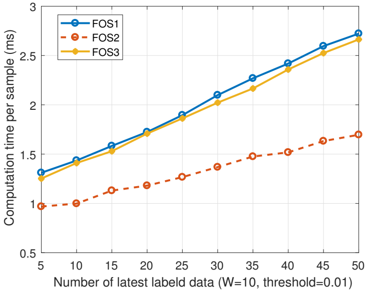


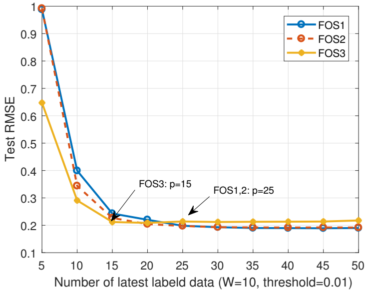
- number of local models decrease among the increase of W
- the test RMSE increase among the increase of W

window size

a small window size W can improve the ability of capturing local characteristics, obtained at large number of local models, and vice verse.

Parameter sensitive analysis



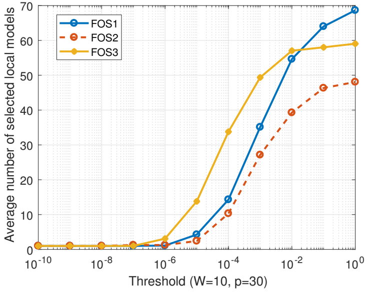


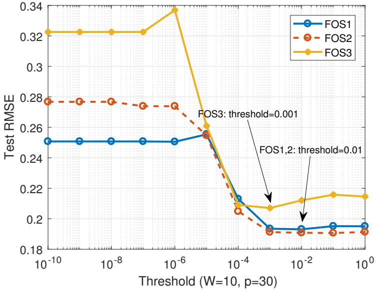
- computation time increase among the increase of p
- the test RMSE decrease among the increase of p

innovation length

a large p can improve the model robustness against noise, but obtained at heavy computation, and vice verse.

Parameter sensitive analysis





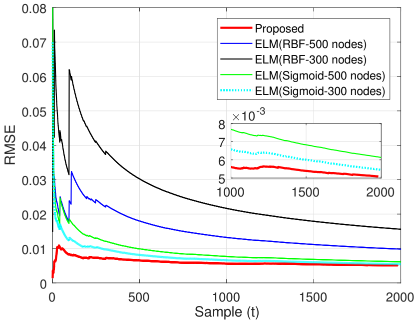
- number of selected local models increase among the increase of ε
- the test RMSE decrease among the increase of ε

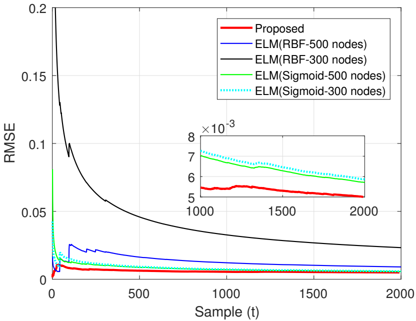
threshold

a large ε can improve the modeling accuracy with more local models combined, thus increasing the computation time, and vice verse.

Comparative study

online RMSE learning curves of three models:
ELM (RBF nodes), **ELM (Sigmoid nodes)**, **Proposed**





Comparative study

Table: Comparison of online prediction and adaptive modeling performance for the OS-ELM and the proposed multiple local model learning approach (ACTpS: Averaged computation time per sample)

Model	RMSE	MAE	ACTpS (ms)	Models
OS-ELM (sigmoid)	2.9911	0.2058	0.18	100
	0.2520	0.1427	1.39	300
	0.2370	0.1427	**5.98**	**500**
OS-ELM (RBF)	1.3952	0.3476	0.56	100
	0.9209	0.2006	2.98	300
	0.4353	0.1676	10.69	500
Proposed	**0.1953**	0.1411	**0.66**	**16-34**

Conclusions

- Our proposed multiple local model learning approach automatically identifies the newly emerging process state during online operation and fits a local linear model to the newly identified process state.
- Adaptive modeling is achieved by a selective ensemble strategy which selects a number of best local linear models from the local model set and optimally combines them to produce the online prediction.
- In the application to a microwave heating system, our proposed approach has been demonstrated to be capable of fast tracking the nonlinear and time-varying characteristics of the underlying system.