Non-Iterative Methods for Classification, Forecasting and Visual Tracking

Part III: Time Series Forecasting

Dr P. N. Suganthan epnsugan@ntu.edu.sg
School of EEE, NTU, Singapore

Some Software Resources Available from:
https://github.com/P-N-Suganthan

DeepLearn 2019
Warsaw, Poland
22 July – 26 July 2019

SEMCCO 2020 & FANCCO 2020

(includes all neural, fuzzy, swarm and evolutionary topics)

HUST, Wuhan, China

June 2020

IEEE SSCI 2019, CIEL 2019, SDE 2019, SIS 2019, etc.

http://ssci2019.org/

(includes all neural, fuzzy, swarm and evolutionary topics)

Randomization-Based ANN, Pseudo-Inverse Based Solutions, Kernel Ridge Regression, Random Forest and Related Topics

http://www.ntu.edu.sg/home/epnsugan/index_files/RNN-Moore-Penrose.htm
http://www.ntu.edu.sg/home/epnsugan/index_files/publications.htm

NANYANG TECHNOLOGICAL UNIVERSITY

Outline

Part I: Classification by

- Neural Networks such as the RVFL
- Kernel Ridge Regression (KRR)
- Random Forest (RF)

Part II: Time Series Forecasting & Classification

(by RVFL, RF, Deep Learners and others)

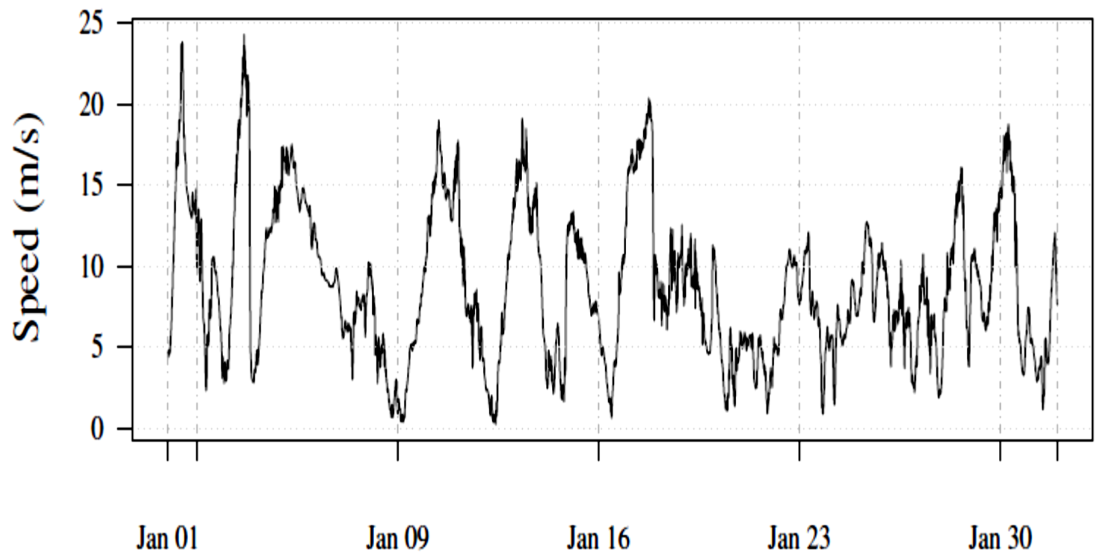
Part III: Visual Tracking

(by RVFL, RF, KRR, etc.)

Time Series Forecasting

- Introduction
- Data Preprocessing
- Time Series Forecasting with Oblique Random Forest
- RVFL for Forecasting
- EMD based Random Vector Functional Link network & other variants
- Ensemble Incremental Learning with Random Vector Functional Link Network
- Ensemble Deep Learning for Time Series Forecasting
- Empirical Mode Decomposition based Ensemble Deep Learning
- Incremental RVFL based Crude Oil Price Forecasting

Introduction – Time Series



A **time series** is a sequence of data points that

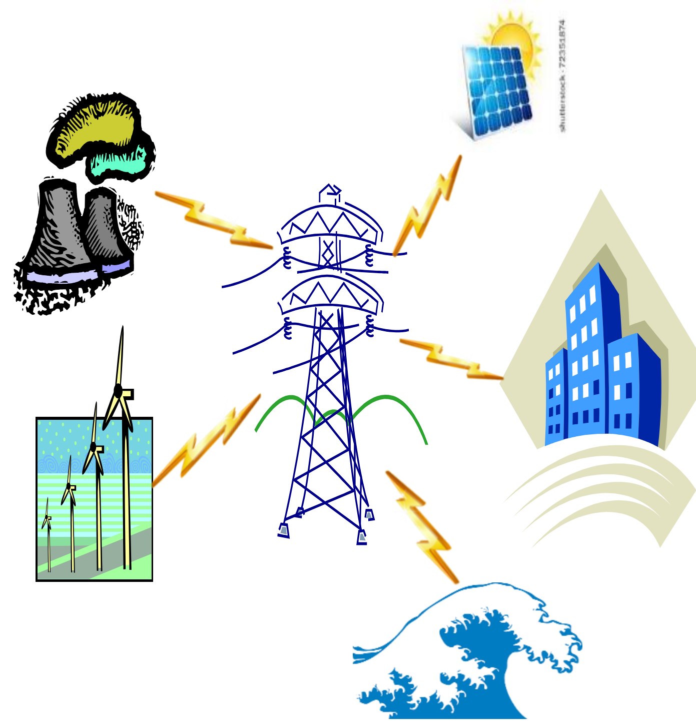
1. Consists of successive measurements made over a time interval
2. The time interval is continuous
3. The distance in this time interval between any two consecutive data Point is the same
4. Each time unit in the time interval has at most one data point
5. Univariate (working with a single series) / multivariate (multiple series)
6. Classification or forecasting.

NANYANG TECHNOLOGICAL UNIVERSITY

Introduction – Applications

Finance

Energy



6

Time Series Forecasting

- Historical value => future value
- Current observations => future value
- Hybrid of the above two

- Statistical approach
- Physical approach based on differential eqns.
- Computational intelligence based approach
- Hybrid/ensemble approach

Y. Ren, L. Zhang, and P. N. Suganthan, “Ensemble Classification and Regression – Recent Developments, Applications and Future Directions,” IEEE Comput. Intell. Mag., 2016, doi: 10.1109/MCI.2015.2471235

Time Series Forecasting (Cont'd)

- Wind speed/power forecasting
- Electricity Load demand forecasting
- Electricity Price forecasting
- Solar irradiance/power forecasting
- Load/wind/solar power ramp forecasting
- Stock price
- Exchange rates, etc.

Time Series Forecasting

- Introduction
- Data Preprocessing
- Time Series Forecasting with Oblique Random Forest
- RVFL for Forecasting
- EMD based Random Vector Functional Link network & other variants
- Ensemble Incremental Learning with Random Vector Functional Link Network
- Ensemble Deep Learning for Time Series Forecasting
- Empirical Mode Decomposition based Ensemble Deep Learning
- Incremental RVFL based Crude Oil Price Forecasting

NANYANG TECHNOLOGICAL UNIVERSITY

Data Preprocessing

- The presentation and quality of training data is important

- Data preprocessing includes:
  - Data cleaning
  - Normalization
  - Feature extraction and selection

Data Preprocessing

- Data cleaning
  - Outlier
  - Incomplete data
  - White noise

- Normalization
  - Min-max normalization

$$v' = \frac{v - min_A}{max_A - min_A}(newmax_A - newmin_A) + newmin_A$$

  - Z-score normalization

$$v' = \frac{v - mean_A}{std_A}$$

NANYANG TECHNOLOGICAL UNIVERSITY

Data Preprocessing

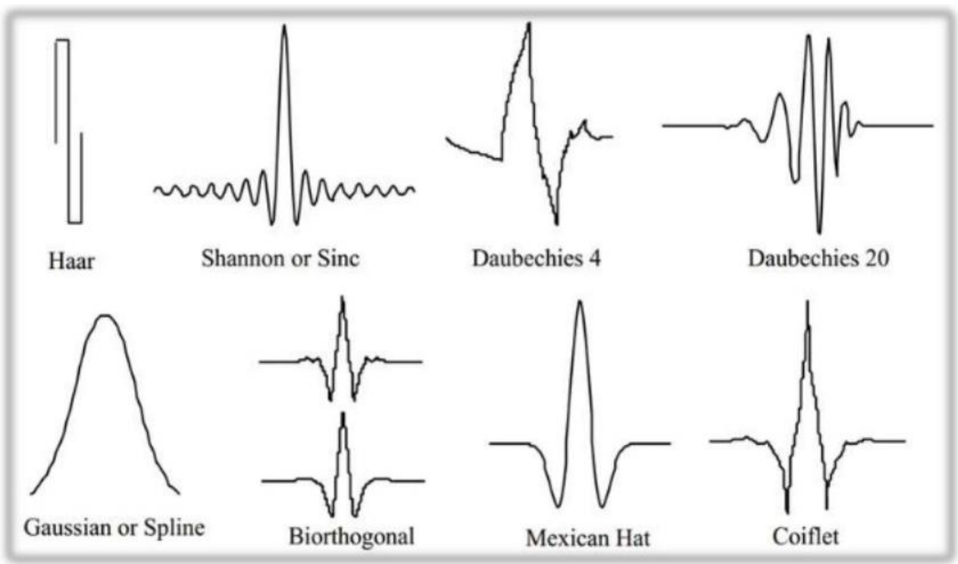
- Feature extraction: a process to construct new feature from the basic feature set
  - Principal Component Analysis
  - Multidimensional Scaling
  - Extraction of local features

- Empirical Mode Decomposition & Discrete Wavelet Transform
  - Decompose load demand time series for feature extraction
  - Remove the white noise

Discrete Wavelet Transform

- Wavelet transform decomposes a signal into a set of basis functions, which are called wavelets.
- Steps for Discrete wavelet transform representation:
  1. Take a mother wavelet and compare it to section at the start of the original signal, and calculate a correlation coefficient C
  2. Shift the wavelet to the right and repeat step 1 until the whole signal is covered
  3. Scale the wavelet and repeat steps 1 through 2.
  4. Repeat steps 1 through 3 for all scales.

Examples for Mother wavelets



Picture source: https://www.slideshare.net/RajEndiran1/introduction-to-wavelet-transform-51504915/18

Empirical Mode Decomposition (EMD)

1. Identify all local maxima and local minima in the TS $\mathbf{x}(t)$ and interpolate all local extrema to form an upper envelope and a lower envelope $u(t)$ and $l(t)$, respectively.

2. Find the mean of upper and lower envelopes $m(t) = \frac{u(t)+l(t)}{2}$.

3. Subtract the mean from the original TS to obtain a detailed component $d(t) = x(t) - m(t)$.

4. If $m(t)$ and $d(t)$ satisfy one of the stopping criterion, then the first IMF $c_1(t) = m(t)$ and the first residue $r_1(t) = d(t)$. The stopping criteria are: (i) $m(t)$ approaches zero, (ii) the number of local extrema and the number of zero-crossings of $d(t)$ differs at most by one or (iii) the user-defined maximum iteration is reached.

5. Else, repeat steps 1 to 4 for $d(t)$ until $c_1(t)$ and $r_1(t)$ are obtained. These iterative steps are known as *sifting*.

6. For $r_1(t)$, repeat Steps 1 to 5 until all the IMFs and the residue are obtained.

Finally:

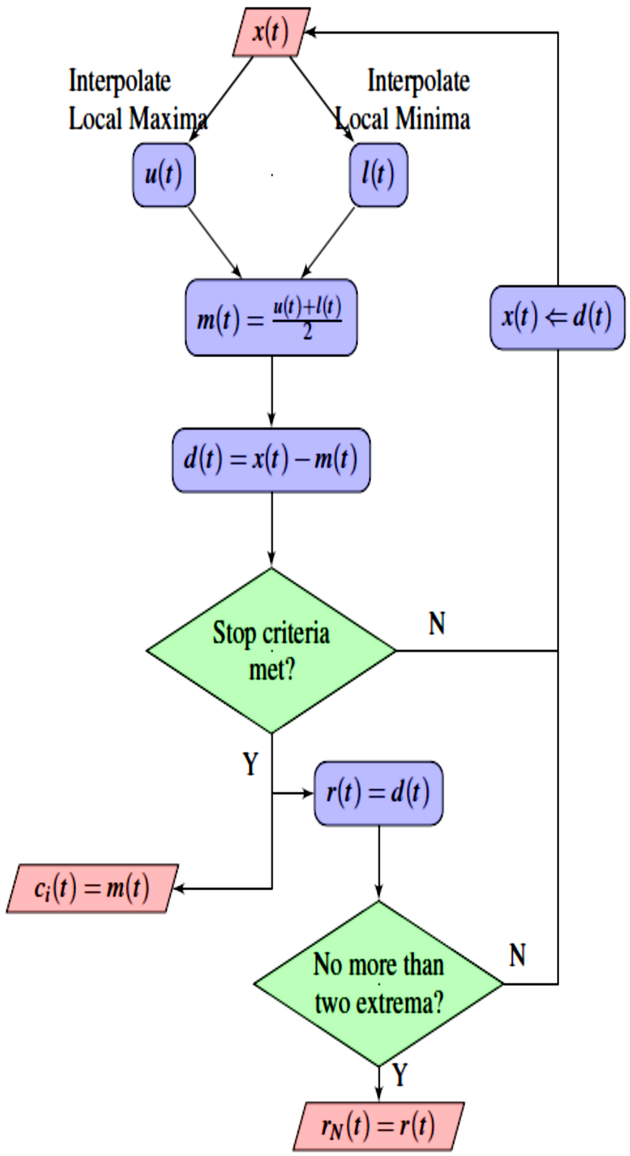
$$x(t) = \sum_{i=1}^{N} c_i(t) + r_N(t)$$

IMF: Intrinsic Mode Functions

N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, “The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis,” Proc. Royal Society London A, vol. 454, pp. 903–995, 1998.

NANYANG TECHNOLOGICAL UNIVERSITY

Empirical Mode Decomposition (EMD)



Finally:

$$x(t) = \sum_{i=1}^{N} c_i(t) + r_N(t)$$

N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, “The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis,” Proc. Royal Society London A, vol. 454, pp. 903–995, 1998.

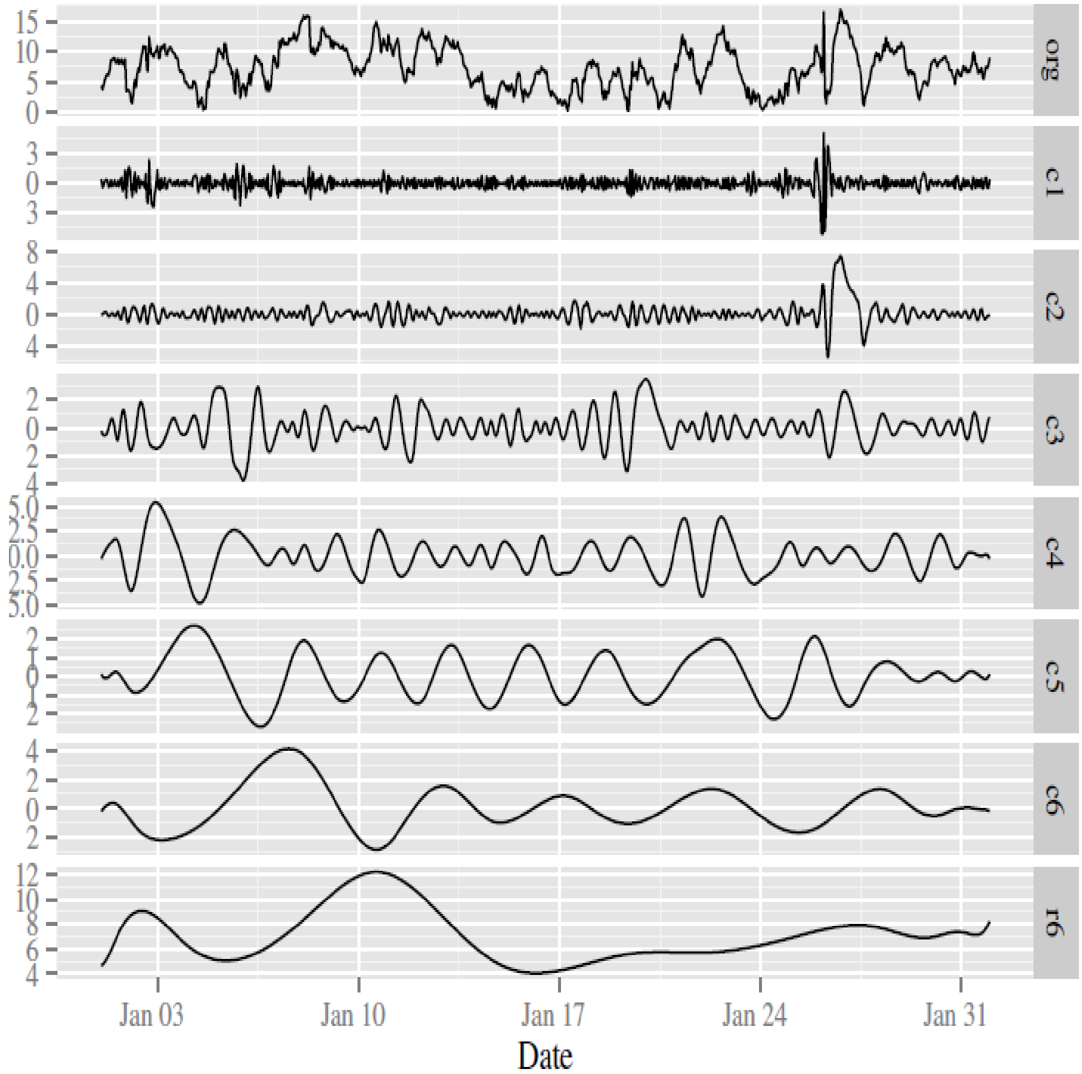
Empirical Mode Decomposition (EMD) (cont'd)

- Adaptive, Local, Orthogonal, Completeness
- Decompose complex time series into simpler time series (narrow band, symmetric)
- Reveal hidden features/correlations of the time series

- Mode mixing problem:
  - a sub series consists of signal spanning a wide band of frequency or
  - more than one sub series contain signals in a similar frequency band
- Ensemble of EMD => solve the problem

G. Rilling, P. Flandrin, and P. Gonc¸alv`es, "On empirical mode decomposition and its algorithms," in Proc. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP'03), no. 3, Grado, Italy, 2003, pp. 8–11.

Empirical Mode Decomposition (EMD) (cont'd)



Ensemble of EMD

- Ensemble EMD (EEMD)
  - Add uncorrelated Gaussian noise to the original time series,
  - Repeat EMD to the noise added series
  - Combine the results: noise will cancel each other
  - But completeness is violated

1. Create a collection of noise added original TS: $x^i(t) = x(t) + \varepsilon^i(t),\ i \in \{1,\ldots,I\}$, where $\varepsilon(t)$ are independent Gaussian white noise.

2. For each $x^i(t)$, apply EMD (discussed in Section 4.1) to obtain the decomposed IMFs and residue: $x^i(t) = \sum_{j=1}^{N} c^i_j + r^i_N$.

3. In order to reconstruct back the original TS, one just needs to average on all trials:

$$x(t) = \frac{1}{I}\left(\sum_{i=1}^{I}\sum_{j=1}^{N} c^i_j + r^i_N\right) + \varepsilon_I \qquad (5.1)$$

where $\varepsilon_I = \frac{\varepsilon}{\sqrt{I}}$ [149] is the aggregated error introduced by the white noise.

Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” Advances in Adaptive Data Analysis, vol. 1, no. 1, pp. 1–41, 2009.

Ensemble of EMD (cont'd)

- Complementary EEMD (CEEMD)
  - Gaussian noise are in pair and complementary to each other
  - Completeness is retained
  - Needs more trials (ensembles)

$$\varepsilon^i(t) \in \{\varepsilon_+^{i/2}(t), \varepsilon_-^{i/2}(t)\}$$

$$\text{where } \varepsilon_+^{i/2}(t) + \varepsilon_-^{i/2}(t) = 0,\ i \in \{1, \ldots, I\}.$$

J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complimentary ensemble empirical mode decomposition: a novel noise enhanced data analysis method," Advances in Adaptive Data Analysis, vol. 2, no. 2, pp. 135– 156, 2010.

Ensemble of EMD (cont'd)

- Complete EEMD with Adaptive Noise (CEEMDAN)
  - Adaptive noise
  - Sequential process
  - Reduce number of trials (ensembles)
  - But cannot do parallel computing

1. Create a collection of noise-added original TS: $x^i(t) = x(t) + w_0\varepsilon^i(t)$, $i \in \{1,\ldots,I\}$, where $\varepsilon(t)$ are independent Gaussian white noise with unit variance and w_0 is a noise coefficient.

2. For each $x^i(t)$, apply EMD (discussed in Section 4.1) to obtain the first decomposed IMF and take average: $c_1(t) = \frac{1}{I}\sum_{i=1}^{I} c_1^i$. Then the first residue is $r_1(t) = x(t) - c_1(t)$.

3. Decompose the noise-added residue $r_1 + w_1 E_1(\varepsilon^i(t))$ to obtain the second IMFs.

$$c_2(t) = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1 + w_1 E_1(\varepsilon^i(t))) \tag{5.3}$$

   where $E_j(\cdot)$ is a function to extract the jth IMF decomposed by EMD

4. Repeat for the remaining IMFs until there are no more than two extrema of the residue.

M. Torres, M. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'11), no. 1520-6149, Prague, Czech Republic, 22-27 May 2011.

Study on EMD-SVR/ANN for wind speed forecasting

- Forecasting models:
  - EMD-ANN, EEMD-ANN, CEEMD-ANN, CEEMDAN-ANN,
  - EMD-SVR, EEMD-SVR, CEEMD-SVR, and CEEMDAN-SVR.
- 12 wind speed TS datasets obtained from National Data Buoy Center (NDBC).
- 70% for training and 30% for testing.
- Forecasting horizon: 1, 3 and 5 hours ahead.
- Scaled to (0,1] interval.
- Compare on RMSE/MASE

$$\mathrm{MASE} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{|e_t|}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}\right) = \frac{\sum_{t=1}^{n}|e_t|}{\frac{n}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}$$

- CEEMDAN: Complementary ensemble empirical mode decomposition with adaptive noise.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y)^2}{n}}.$$

Y. Ren, P. N. Suganthan, and N. Srikanth, "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods," IEEE Trans. Sustain. Energy, vol. 6, no. 1, pp. 236--244, Jan. 2015.

Time Series Forecasting

- Introduction
- Data Preprocessing
- Time Series Forecasting with Oblique Random Forest
- RVFL for Forecasting
- EMD based Random Vector Functional Link network & other variants
- Ensemble Incremental Learning with Random Vector Functional Link Network
- Ensemble Deep Learning for Time Series Forecasting
- Empirical Mode Decomposition based Ensemble Deep Learning
- Incremental RVFL based Crude Oil Price Forecasting

Random Forest for TS forecasting

- Same concept as classification but
  - Classification tree => regression tree
  - Majority vote => averaging for the final output
- Important parameters:
  - n_tree: number of bootstrap samples
  - m_try: number of variables tried at each split
- Criteria at split:
  - Residual sum of square:

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2,$$

  - Or minimize within group variance
  - Or maximize between group variance

Time Series Forecasting with Oblique Random Forest

- Published in Information Sciences[1]
- Contributions:
  - we apply the oblique RF in the context of electricity load time series forecasting
  - instead of impurity score based selection method, we alternatively propose to use least-square error minimization in each node of the decision trees to perform data partitioning, instead of the MPSVM in the classification paper.
  - the advantage of the proposed method is demonstrated using eight generic TS datasets and five electricity load demand datasets compared with six benchmark algorithms

1. X. Qiu, L. Zhang, P. N. Suganthan, and G. A. J. Amaratunga, "Oblique random forest ensemble via least square estimation for time series forecasting," in Information Sciences,vol.420, pp.249–262, 2017.
2. Zhang, Le, and Ponnuthurai N. Suganthan. "Oblique decision tree ensemble via multisurface proximal support vector machine." *Cybernetics, IEEE Transactions on* 45.10 (2015): 2165-2176.

Oblique Random Forest

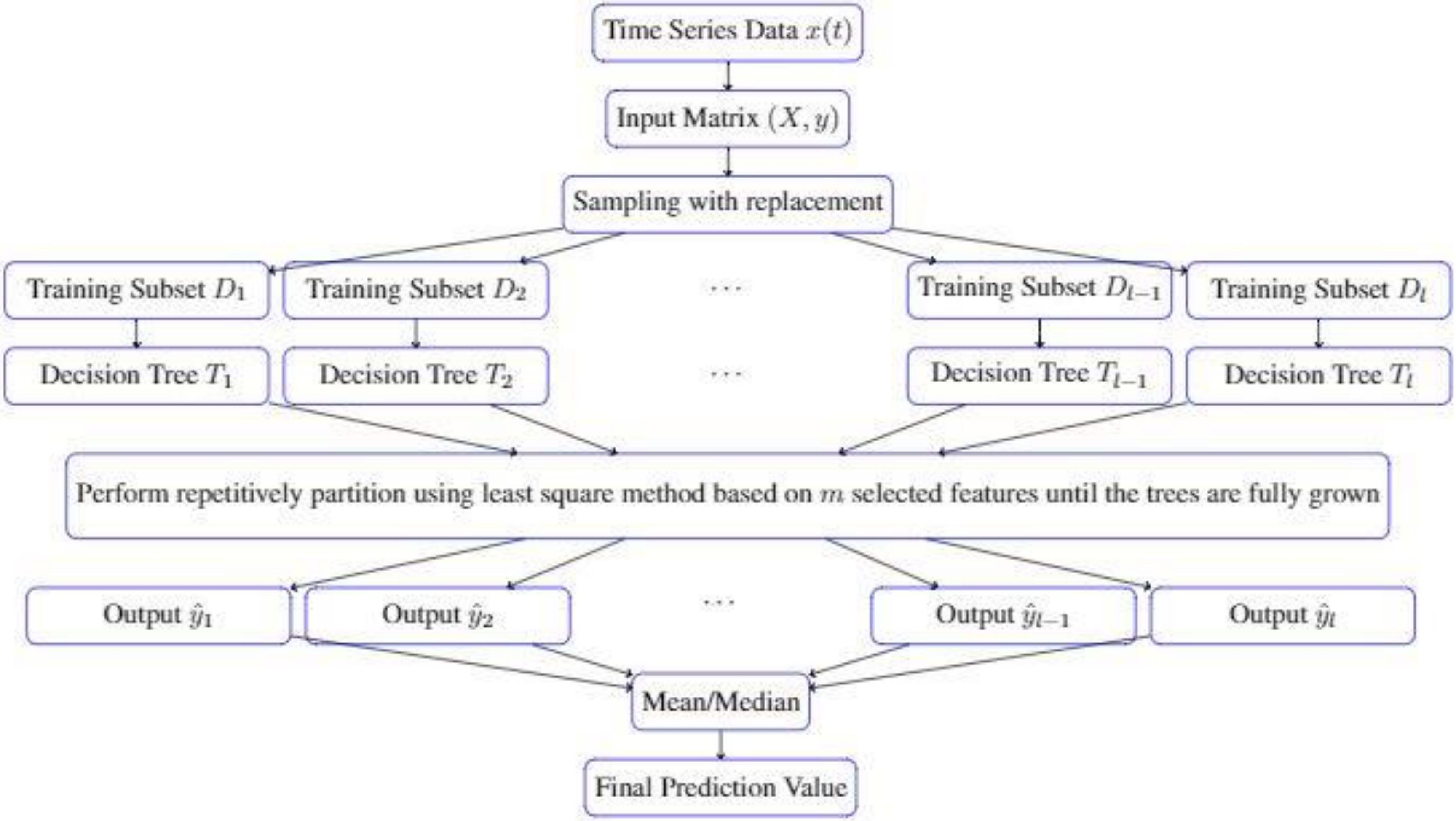


FIGURE 3.3: Schematic Diagram of the Proposed Oblique Random Forest

Computational Complexity

- Orthogonal decision tree:

$$C_{Orth} \approx O(nN(logN)^2)$$

- Oblique decision tree:

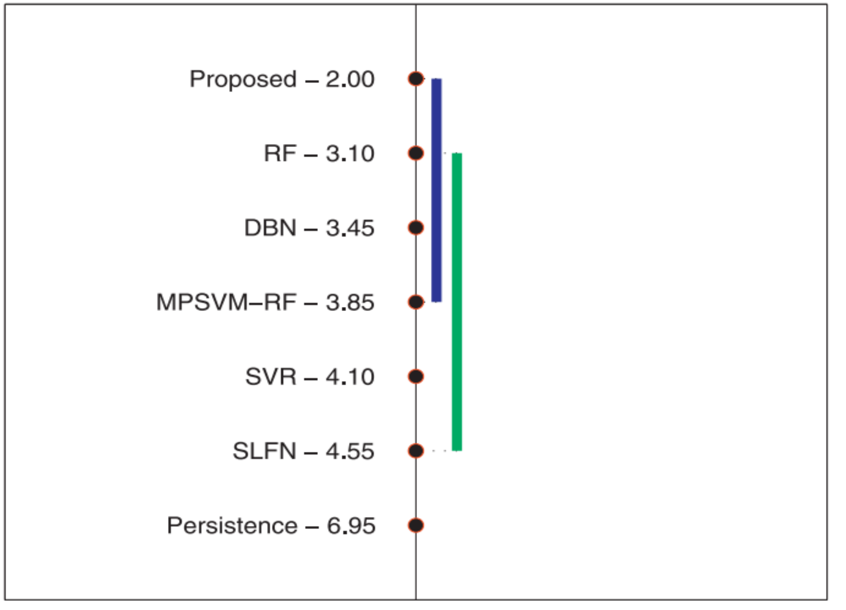
$$C_{Obli} \approx O(Nn^3)$$

- For large datasets:
  - Number of training samples: N
  - Number of features: n
  - $N \gg n \rightarrow C_{Obli} \ll C_{Orth}$

Experiments

- Assessment on generic TS datasets
  - Eight generic TS datasets
  - Compared with five benchmarks

- One day ahead electricity load forecasting
  - Oblique RF has higher rank than RF and MPSVM-RF for short term load forecasting, but statistically they still perform similar since the distance is less than 2.0
  - Decision Tree ensembles perform better than SVR and SLFN

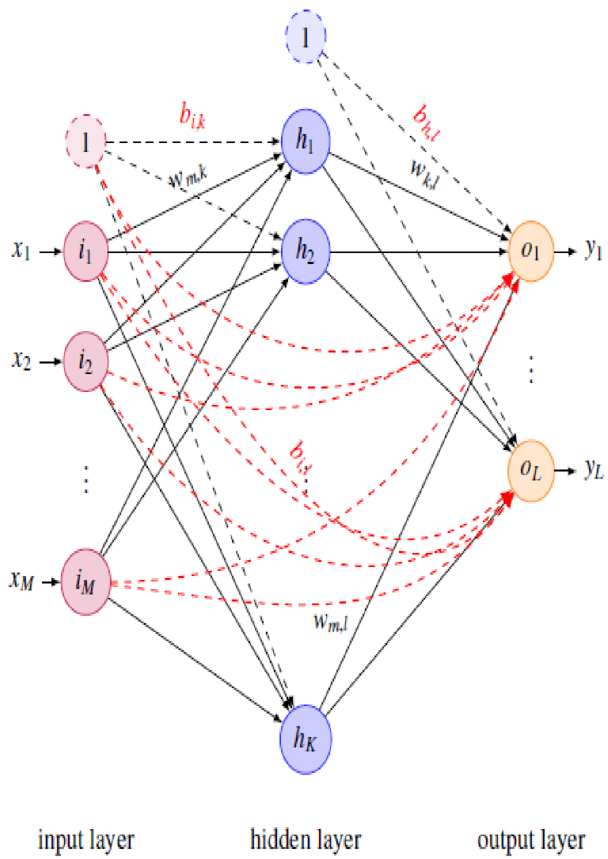


Nemenyi test result based on RMSE. The critical distance is 2.0

Time Series Forecasting

- Introduction
- Data Preprocessing
- Time Series Forecasting with Oblique Random Forest
- RVFL for Forecasting
- EMD based Random Vector Functional Link network & other variants
- Ensemble Incremental Learning with Random Vector Functional Link Network
- Ensemble Deep Learning for Time Series Forecasting
- Empirical Mode Decomposition based Ensemble Deep Learning
- SSCI2018: Incremental RVFL based Crude Oil Price Forecasting

Random vector functional link (RVFL) Network



- No need BP to train the network
- Random weights for $w_{m,k}$
- Least square estimation to obtain $w_{k,l}$ and $w_{m,l}$

1) Assign random values to input weights $w_{i,h}$. The random values are uniformly distributed in the interval $[0,1]$.
2) Obtain the hidden perceptron outputs $\mathbf{A} = \text{logsig}(\mathbf{W}_{I,H} \cdot \mathbf{X})$, where $\mathbf{X}$ is the training data.
3) Apply least square estimation to calculate the output weights $w_{h,o}$ and the direct link weights $w_{i,o}$: $\mathbf{W}_O = (A^T A)^{-1} A^T \mathbf{Y}$, where $\mathbf{W}_O$ is the aggregation of output weights and the direct link weights, and $\mathbf{Y}$ is the training target.

Y.-H. Pao, G.-H. Park, D. J. Sobajic, Learning and generalization characteristics of the random vector functional-link net, Neurocomputing 6 (2) (1994) 163–180.

RVFL variations

- Effects on:
  - Input layer bias
  - Hidden layer bias
  - Direct input-output connections

Method	Input Layer Bias	Hidden Layer Bias	Input-Output Connection	Formula
M1	√	√	√	$h_k = f(\sum_{m=1}^{M} w_{m,k} i_m + b_{i,k})$ $o_l = \sum_{k=1}^{K} w_{k,l} h_k + b_{h,l} + \sum_{m=1}^{M} w_{m,l} i_m + b_{i,l}$
M2 [31]	√	√	×	$h_k = f(\sum_{m=1}^{M} w_{m,k} i_m + b_{i,k})$ $o_l = \sum_{k=1}^{K} w_{k,l} h_k + b_{h,l}$
M3 [27, 6]	√	×	√	$h_k = f(\sum_{m=1}^{M} w_{m,k} i_m + b_{i,k})$ $o_l = \sum_{k=1}^{K} w_{k,l} h_k + \sum_{m=1}^{M} w_{m,l} i_m + b_{i,l}$
M4	√	×	×	$h_k = f(\sum_{m=1}^{M} w_{m,k} i_m + b_{i,k})$ $o_l = \sum_{k=1}^{K} w_{k,l} h_k$
M5	×	√	√	$h_k = f(\sum_{m=1}^{M} w_{m,k} i_m)$ $o_l = \sum_{k=1}^{K} w_{k,l} h_k + b_{h,l} + \sum_{m=1}^{M} w_{m,l} i_m$
M6	×	√	×	$h_k = f(\sum_{m=1}^{M} w_{m,k} i_m)$ $o_l = \sum_{k=1}^{K} w_{k,l} h_k + b_{h,l}$
M7	×	×	√	$h_k = f(\sum_{m=1}^{M} w_{m,k} i_m)$ $o_l = \sum_{k=1}^{K} w_{k,l} h_k + \sum_{m=1}^{M} w_{m,l} i_m$
M8	×	×	×	$h_k = f(\sum_{m=1}^{M} w_{m,k} i_m)$ $o_l = \sum_{k=1}^{K} w_{k,l} h_k, \forall k \in \{1, \ldots, K\}, \forall l \in \{1, \ldots, L\}$

Y. Ren, P. N. Suganthan, N. Srikanth and G. Amaratunga, "Single Hidden Layer Neural Networks with Random Weights for Short-term Electricity Load Demand Forecasting," Information Sciences, 2016.
Schmidt W F, Kraaijveld M A, Duin R P W. Feedforward neural networks with random weights[C]//Pattern Recognition, 1992. Vol. II. Conference B: Pattern Recognition Methodology and Systems, Proceedings., 11th IAPR International Conference on. IEEE, 1992: 1-4.

NANYANG TECHNOLOGICAL UNIVERSITY

RVFL variations (cont'd)

Input layer bias

Comparison	Horizon (h)												
	1	2	4	6	8	10	12	14	16	18	20	22	24
M1 v.s. M5	0.351	0.274	0.531	0.144	0.065	0.184	0.337	0.351	0.204	0.456	0.513	0.862	0.329
M2 v.s. M6	0.421	0.172	0.974	0.492	0.418	0.116	0.255	0.368	0.248	0.443	0.234	0.059	0.173
M3 v.s. M7	0.977	0.517	0.873	0.955	0.05	0.054	0.067	**0.008**	**0.008**	**0.033**	**0.032**	0.126	0.112
M4 v.s. M8	0.339	0.646	0.98	0.907	0.711	0.163	0.522	0.6	0.545	0.6	0.433	0.068	0.315

Hidden layer bias

Comparison	Horizon (h)												
	1	2	4	6	8	10	12	14	16	18	20	22	24
M1 v.s. M3	0.376	0.174	0.249	0.772	0.387	0.186	0.827	0.671	0.619	0.291	0.386	0.968	0.061
M2 v.s. M4	**0.032**	0.236	0.609	0.468	0.176	0.069	**0.013**	**0.02**	**0.023**	0.499	0.352	0.181	0.357
M5 v.s. M7	0.107	**0.036**	0.86	0.407	0.582	0.22	0.122	**0.007**	**0.008**	0.042	**0.037**	0.09	0.237
M6 v.s. M8	0.244	0.977	0.934	0.083	0.254	**0.041**	0.364	0.28	0.158	0.249	0.101	0.063	0.963

Input output Link:

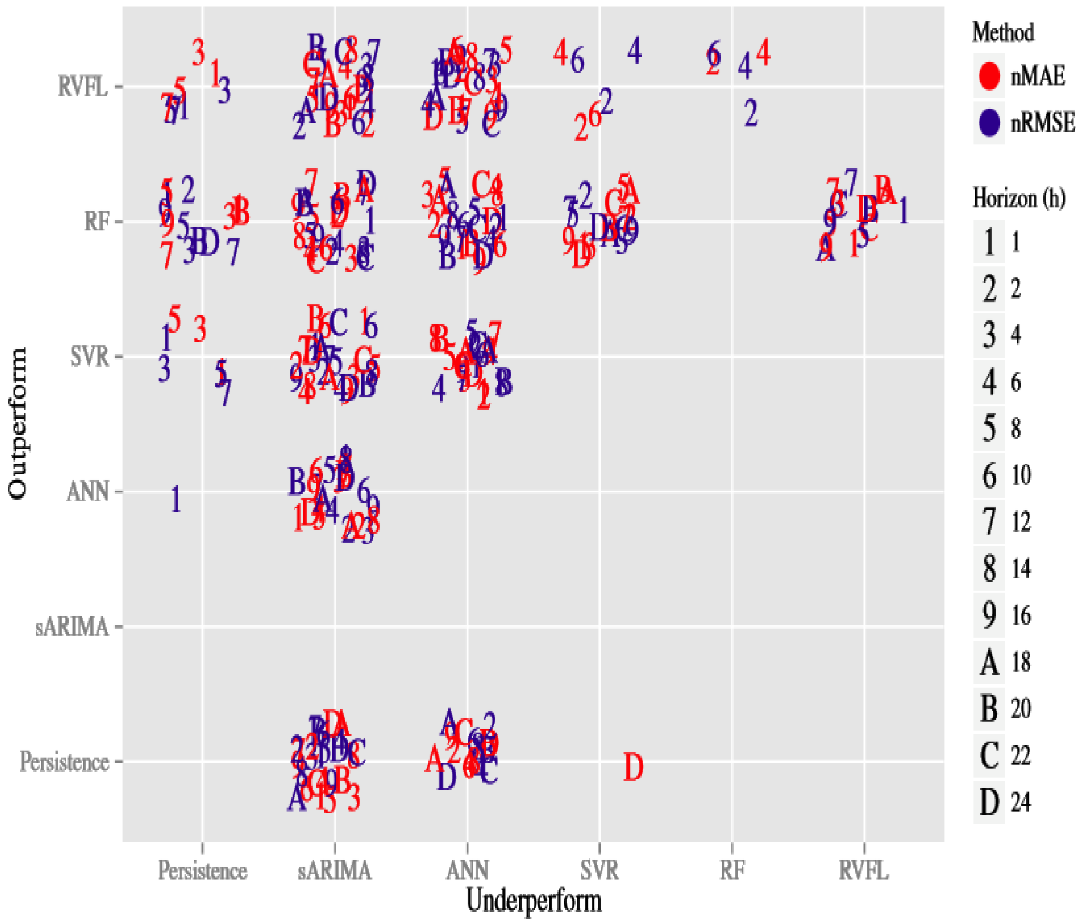
RVFL with direct links out-performed RVFL without direct link like ELM.

Comparison	Horizon (h)												
	1	2	4	6	8	10	12	14	16	18	20	22	24
M1 v.s. M2	3.4e-57	1.2e-29	0.00011	0.13	0.16	0.062	0.025	0.0018	0.00065	0.00026	2.5e-05	3.3e-06	5.7e-09
M3 v.s. M4	1.8e-57	3.3e-31	9.2e-05	0.18	0.22	0.14	0.056	0.0037	0.0011	0.00058	4.8e-05	3.9e-06	7e-10
M5 v.s. M6	4.2e-56	1.5e-32	0.0019	0.22	0.76	0.86	0.61	0.23	0.12	0.017	0.00023	8.3e-05	4.4e-06
M7 v.s. M8	1e-57	2.4e-39	6.3e-06	0.046	0.49	0.72	0.57	0.48	0.32	0.083	0.00057	0.00012	3.3e-08

RVFL for load demand forecasting

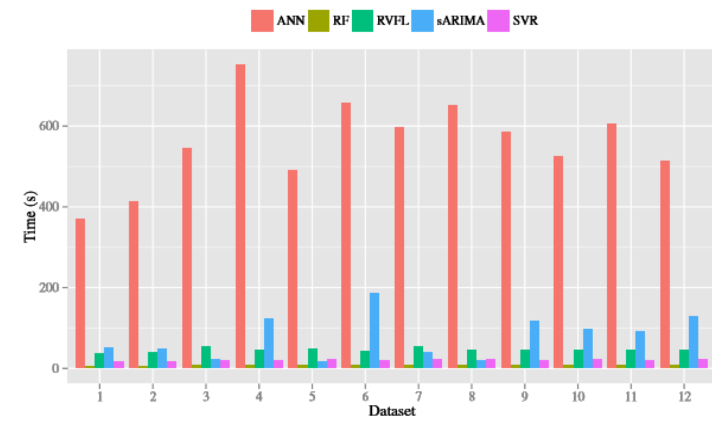
- Input layer bias and hidden layer bias insignificantly affected the performance
- whereas the direct input– output connections significantly improved the performance
- Quantile scaling algorithm has improved the performance for 1–4 hour and 18–24 hour ahead forecasting horizons.
- Feature selection based on partial auto correlation function or seasonal auto-regression has consistently degraded the performance on the seasonal time series.

RVFL for load demand forecasting (cont'd)

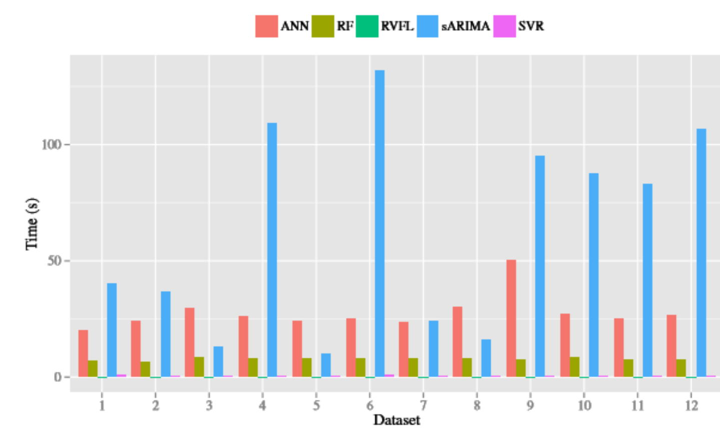


Performance

Training time



Testing time



NANYANG TECHNOLOGICAL UNIVERSITY

RVFL for load demand forecasting (cont'd)

- No clear overall advantage of input layer and hidden layer biases. However, the input layer biases are necessary for the neural networks to function properly as a universal approximator => recommend retaining biases as the selection choices as they may be beneficial for some forecasting problems.
- Compared with reported non-ensemble forecasting methods such as the persistence method, seasonal ARIMA and artificial neural networks, the RVFL network has significantly better performance.
- The RVFL network is underperformed by random forest, which is an ensemble method.
- The computation time (including training and testing) of the RVFL network is the shortest compared with the reported methods.

Conclusions

- RVFL network with input bias, without hidden bias, with direct input—output connections is the best configuration
- RVFL network has significantly better performance than non-ensemble methods for wind speed TS forecasting
- RVFL network has better performance and shorter time for wind power ramp forecasting
- RF is also highly competitive when the original time series data is used (without EMD, DWT, etc.)

Time Series Forecasting

- Introduction
- Data Preprocessing
- Time Series Forecasting with Oblique Random Forest
- RVFL for Forecasting
- EMD based RVFL network & other variants
- Ensemble Incremental Learning with Random Vector Functional Link Network
- Ensemble Deep Learning for Time Series Forecasting
- Empirical Mode Decomposition based Ensemble Deep Learning
- Incremental RVFL based Crude Oil Price Forecasting

Empirical Mode Decomposition based Random Vector Functional Link network

- Deep learning ensembles need long computation time
- Instead of DBN, this work investigates combining fast training non-iterative model RVFL with EMD
- Also focus on electricity load forecasting

X. Qiu, P. N. Suganthan, and G. A. J. Amaratunga, "Electricity Load Demand Time Series Forecasting with Empirical Mode Decomposition based Random Vector Functional Link Network," in: Proc. IEEE Conference on Systems, Man and Cybernetics (SMC2016), Budapest, Hungary, Oct. 2016.

NANYANG TECHNOLOGICAL UNIVERSITY

Experiment Conclusions

- For short-term electricity load demand time series forecasting:
  - EMD based hybrid methods outperform the corresponding single structure models.
  - The computation time of the RVFL network is the shortest among all of the benchmarks.
  - EMD-RVFL has comparable learning capability as deep learning ensembles while has the advantages of high efficiency.

Other EMD based Ensemble Models

- EMD based ensemble kernel machines[1]
  - Focus on short term electricity price forecasting
- EMD based v-support vector regression model[2]
  - Focus on stock price TS forecasting
- EMD based adaboost-BP neural network[3]
  - Focus on wind power forecasting

1. X. Qiu, P. N. Suganthan, and G. A. J. Amaratunga, "Short-term Electricity Price Forecasting with Empirical Mode Decomposition based Ensemble Kernel Machines," Procedia Computer Science, vol. 208, pp. 1308–1317, 2017.
2. X. Qiu, H. Zhu, P. N. Suganthan, and G. A. J. Amaratunga, "Stock Price Forecasting with Empirical Mode Decomposition Based Ensemblev-Support Vector Regression Model," in: Proc. International Conference on Computational Intelligence, Communications, and Business Analytics, Sep. 2017.
3. Y. Ren, X. Qiu, and P. N. Suganthan, "EMD based AdaBoost-BPNN method for wind speed forecasting," in: Proc. IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL'14), Orlando, US, Dec. 2014.

Time Series Forecasting

- Introduction
- Data Preprocessing
- Time Series Forecasting with Oblique Random Forest
- Random Vector Functional Link (RVFL) for Forecasting
- EMD based RVFL network & other variants
- Ensemble Incremental Learning with RVFL Network
- Ensemble Deep Learning for Time Series Forecasting
- Empirical Mode Decomposition based Ensemble Deep Learning
- Incremental RVFL based Crude Oil Price Forecasting

Ensemble Incremental Learning with Random Vector Functional Link Network

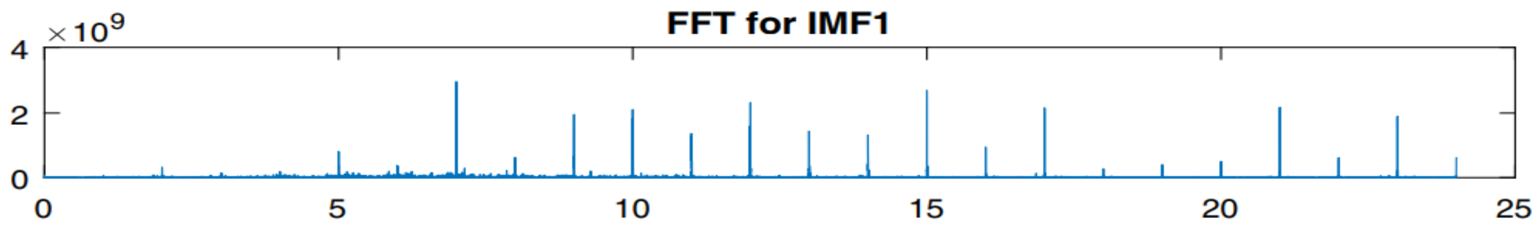
- Contributions:
  - a sequentially combined decomposition method using DWT and EMD is designed for TS signal.
  - accomplish incremental learning with RVFL
  - an incremental ensemble learning approach is proposed for TS forecasting, which is composed of DWT and EMD, along with incremental RVFL network.

X. Qiu, P. N. Suganthan, and G. A. J. Amaratunga, “Ensemble Incremental Learning Random Vector Functional Link Network for Short-term Electricity Load Demand Fore-casting,” Knowledge-based Systems, vol. 145, pp. 182-196, 2018

Sequential Combination of DWT & EMD

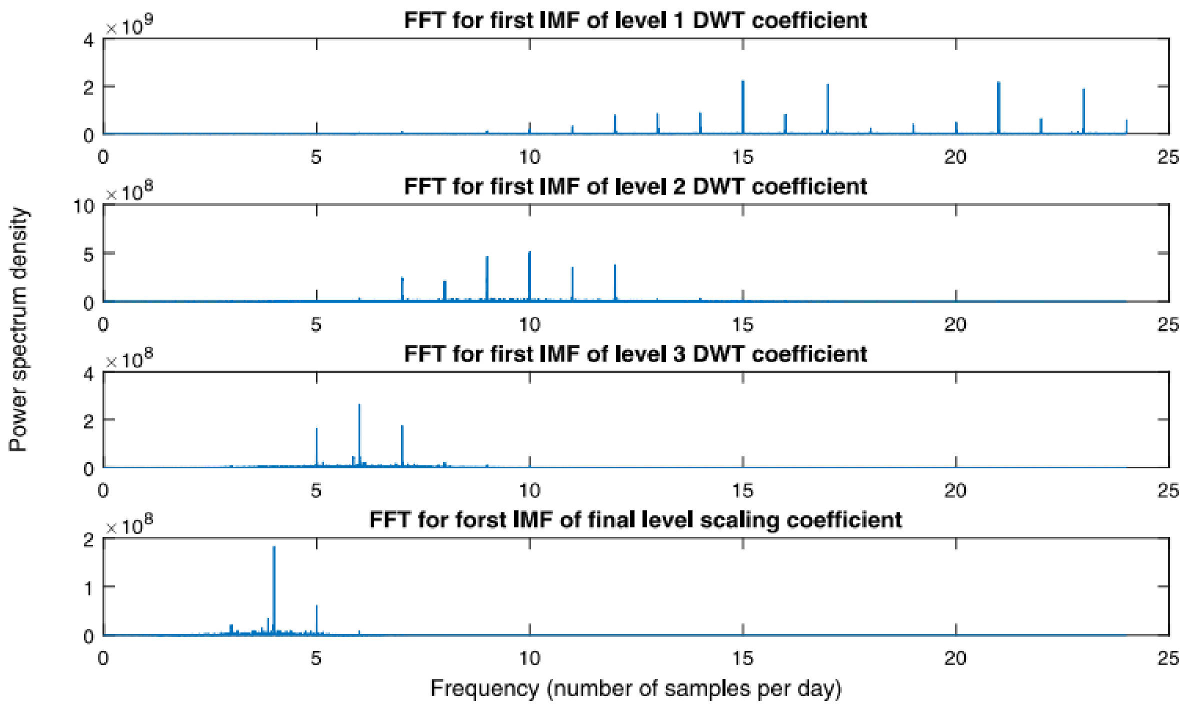
- Mode mixing problem of EMD:
  - One IMF may consist of signal spanning a wide band of frequency, or more than one IMFs contain signals in a similar frequency band
  - Ensemble EMD is one of the solutions
  - Alternatively, Discrete Wavelet Transform can help deal with the frequency distribution issue.

Sequential Combination of DWT & EMD

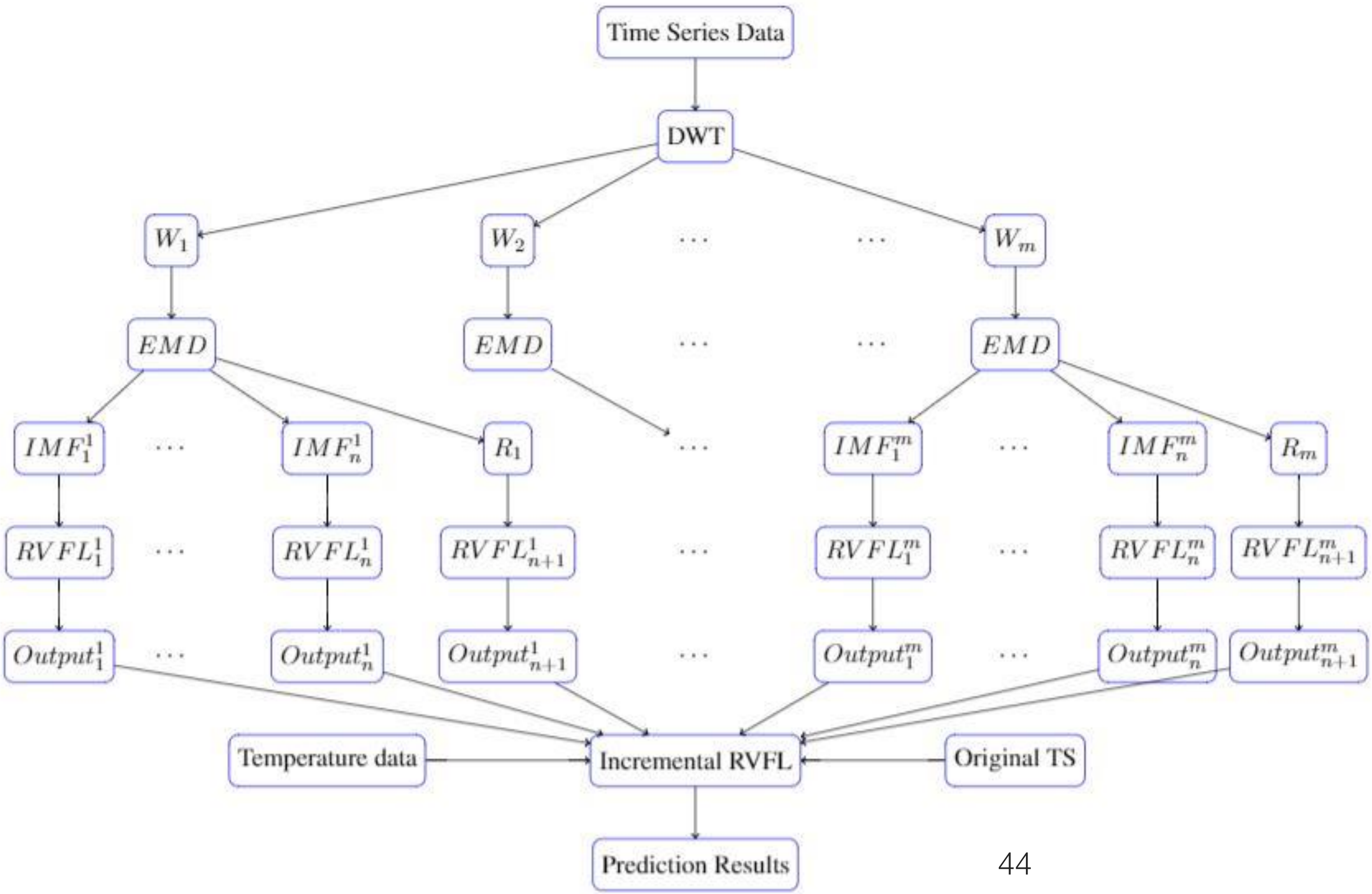


Frequency spectrum for the first IMF obtained from EMD only

Frequency spectrum for the first IMF of each wavelet based sub-signal



DWT-EMD based Incremental RVFL Network

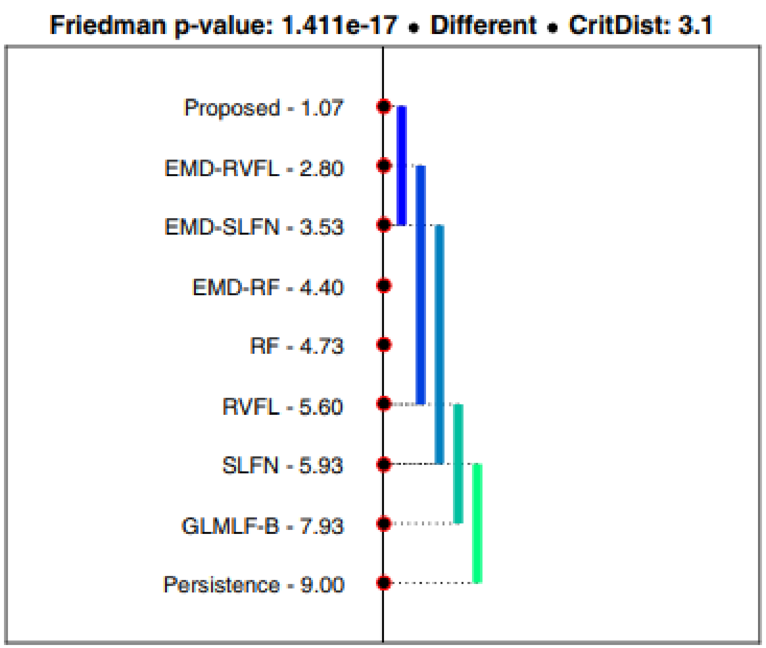


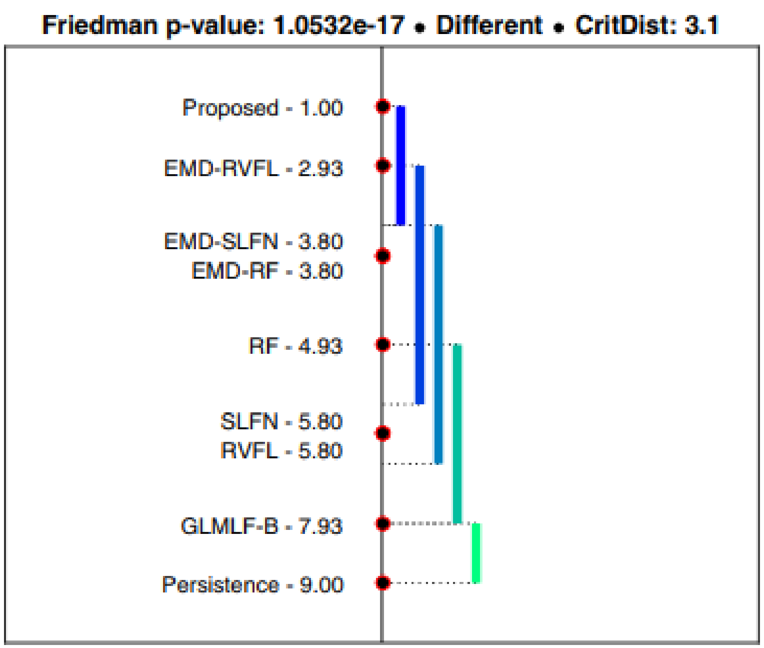
DWT-EMD based Incremental RVFL Network

- Off-line Training Procedures:
  - Apply DWT to decompose original TS into several frequency components
  - Apply EMD to decompose each frequency component into several IMFs and one residue
  - Train an RVFL network for each extracted IMF and residue
  - Aggregate all the outputs by another RVFL to generate the final prediction results.
- Online Updating Procedures:
  - When new input sample occurs, DWT-EMD is re-employed on the signal to get the new input pattern, which can be used to updated the weights of RVFL models.
  - Evaluate the updated model by validation data to determine keep the update or not.

Experiment Results

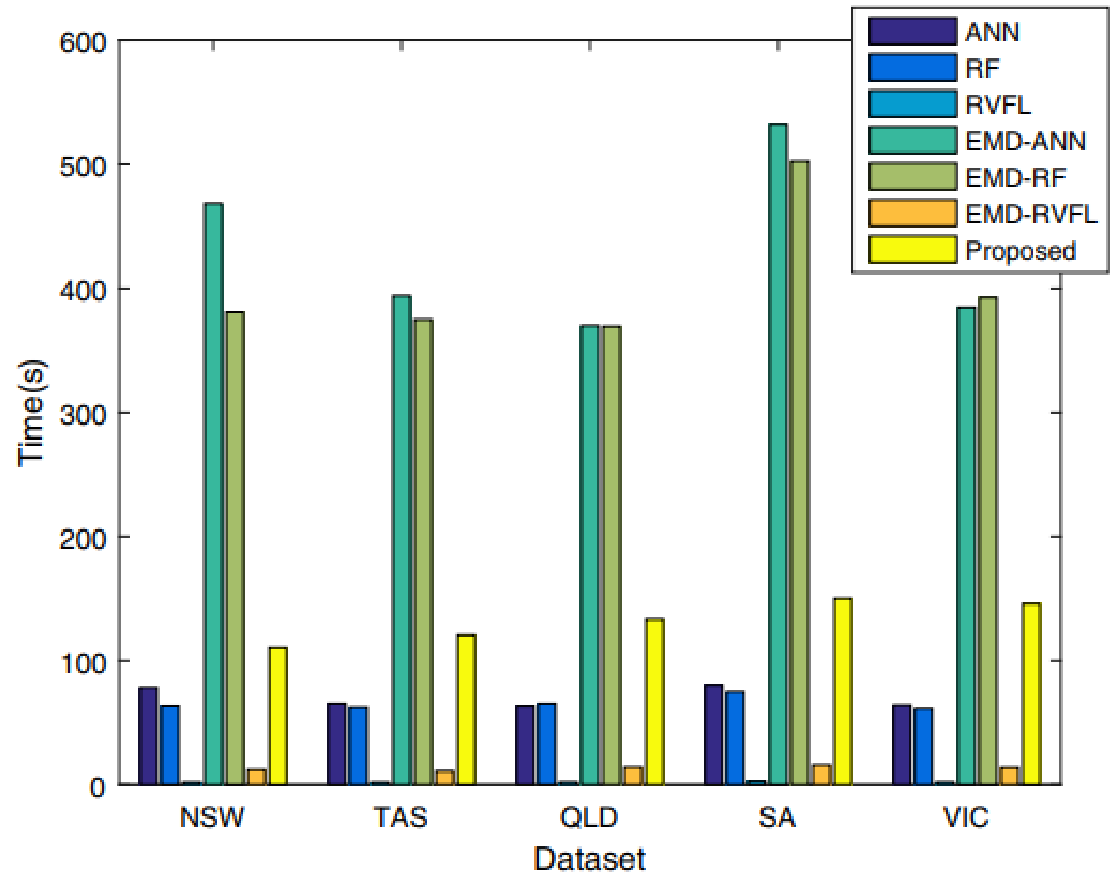
- One-day-ahead electricity load forecasting
  - Compared with RVFL, SLFN, RF and their EMD ensemble models
  - We focus on how to further improve EMD based ensemble models.





Computation Time Comparison

- RVFL is much faster than ANN and RF for medium size datasets
- Thus the proposed DWT-EMD-RVFL model also has a reasonable fast computation speed.

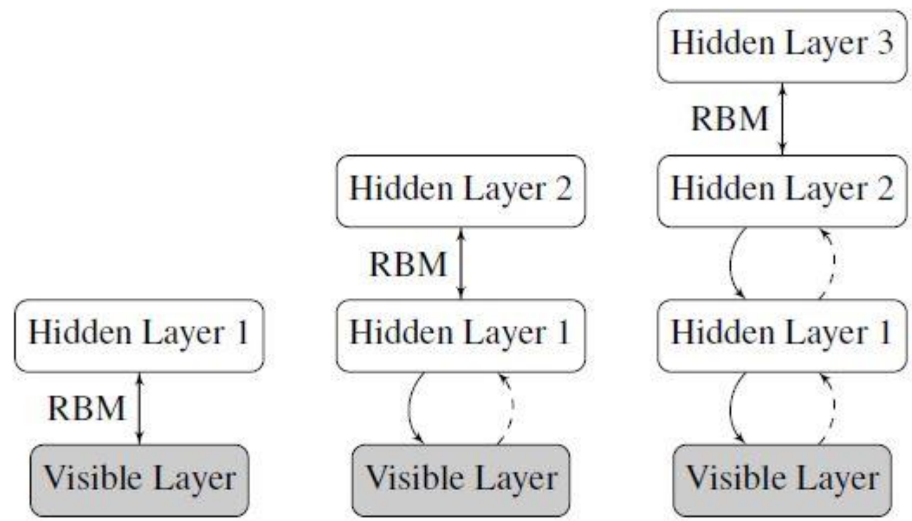


Time Series Forecasting

- Introduction
- Data Preprocessing
- Time Series Forecasting with Oblique Random Forest
- RVFL for Forecasting
- EMD based Random Vector Functional Link network & other variants
- Ensemble Incremental Learning with Random Vector Functional Link Network
- Ensemble Deep Learning for Time Series Forecasting
- Empirical Mode Decomposition based Ensemble Deep Learning
- Incremental RVFL based Crude Oil Price Forecasting

Deep Belief Network

- Layer-wise greedy pre-training algorithm
- Use restricted Boltzmann machines to learn the probability over the input dataset
- Train multiple layers:
  - Train the first layer first and freeze it
  - Use the conditional expectation of the output as input to next layer

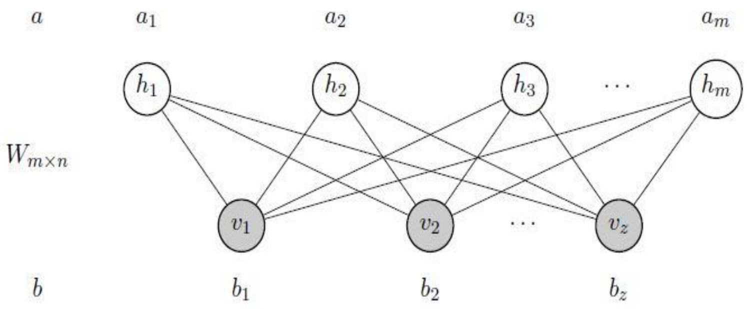


NANYANG TECHNOLOGICAL UNIVERSITY

Restricted Boltzmann Machine

- Joint Probability:

$$P_{h,v}(h,v) = \frac{1}{Z_{h,v}} \cdot e^{(v^TWh+(v-b)^T(v-b)+a^Th)}$$

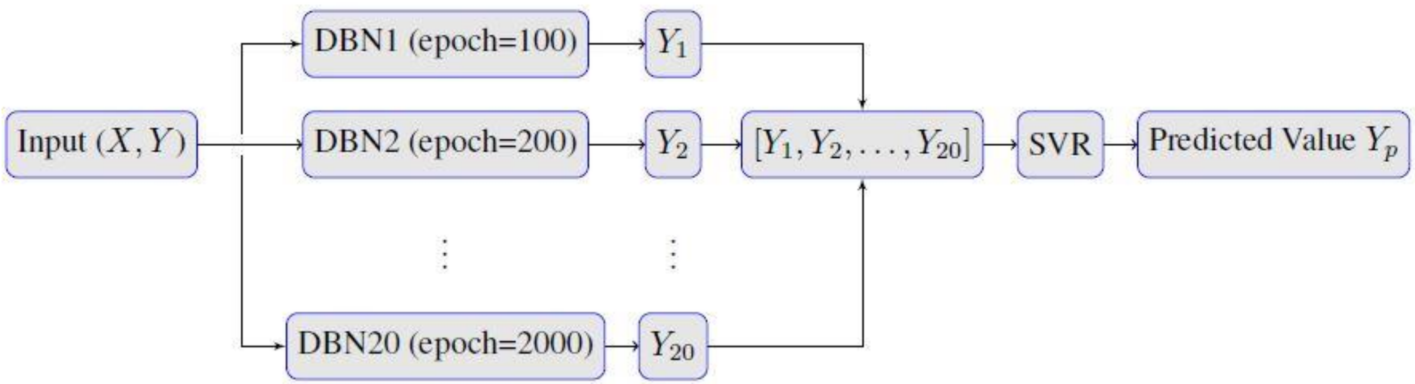


- RBM parameters:
  - Trained in an unsupervised fashion
  - Maximizing the likelihood over training samples
  - Using approximate contrastive divergence algorithm

$$\mathcal{L} = \prod_t \sum_h P_{h,v}(h, v(t))$$

Ensemble Deep Learning for Time Series Forecasting

- Published in 2014 IEEE Symposium Series on Computational Intelligence (SSCI2014)[1]
- Ensemble deep learning algorithm composed of
  - Deep Belief Networks (DBNs) trained using different number of epochs
  - Support Vector Regression (SVR) takes the outputs of the DBNs as its inputs and outputs as the final prediction



- Ensemble deep learning method demonstrates much stronger ability on real complicated problems

1. X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan, and G. Amaratunga, "Ensemble Deep Learning for Regression and Time Series Forecasting," in: Proc. IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL'14), Orlando, US, Dec. 2014. (First ensemble deep learning paper for forecasting)

Time Series Forecasting

- Introduction
- Data Preprocessing
- Time Series Forecasting with Oblique Random Forest
- RVFL for Forecasting
- EMD based Random Vector Functional Link network & other variants
- Ensemble Incremental Learning with Random Vector Functional Link Network
- Ensemble Deep Learning for Time Series Forecasting
- Empirical Mode Decomposition based Ensemble Deep Learning
- Incremental RVFL based Crude Oil Price Forecasting

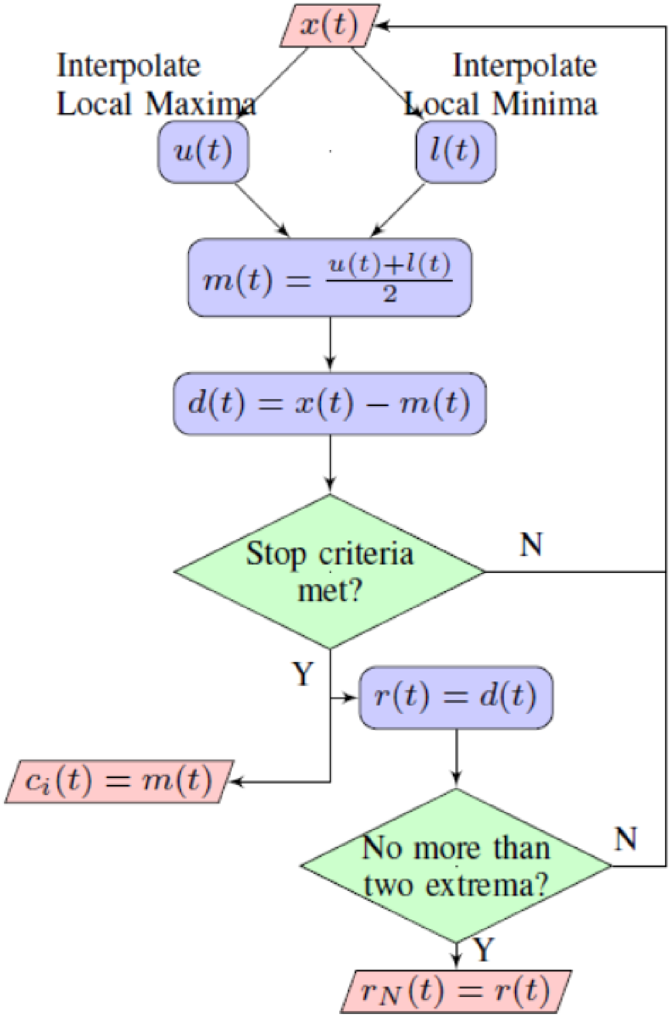
NANYANG TECHNOLOGICAL UNIVERSITY

Empirical Mode Decomposition

- Decompose a signal into several intrinsic mode functions with one residue

- Be effective for non-stationary and nonlinear data sets

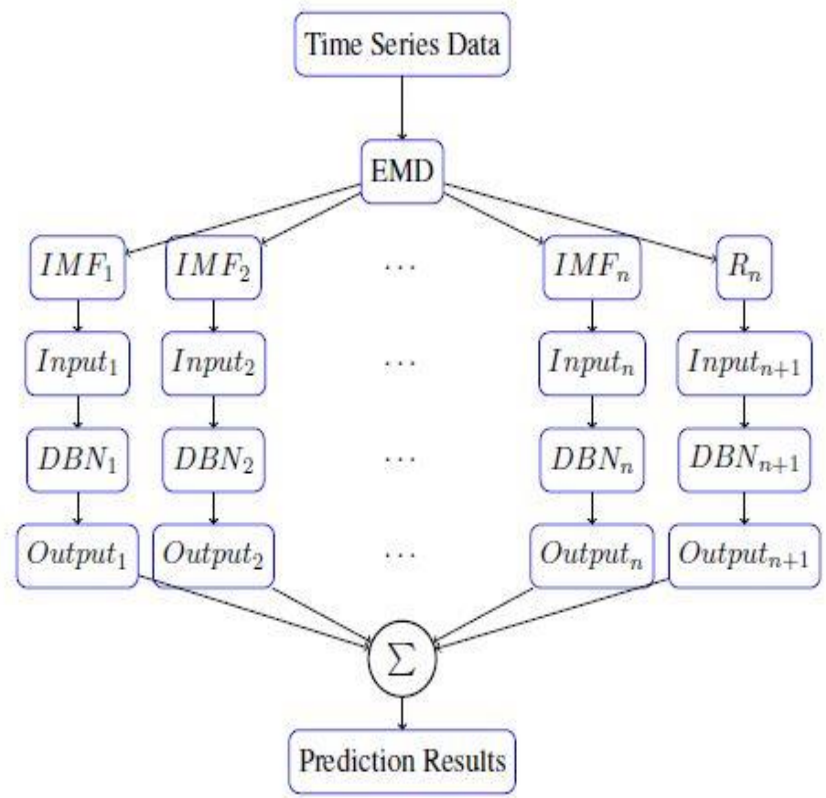
$$x(t) = \sum_{i=1}^{n} c_i + r_n$$



NANYANG TECHNOLOGICAL UNIVERSITY

Empirical Mode Decomposition based Ensemble Deep Learning

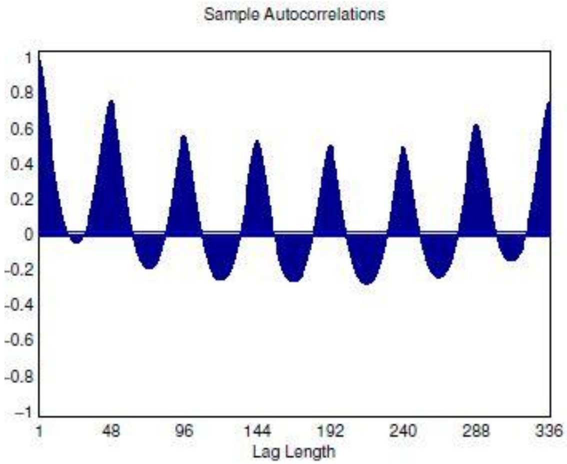
- Use the concept of "divide and conquer"
- Ensemble deep learning algorithm composed of
  - Empirical Mode Decomposition (EMD) decomposes Time Series into several Intrinsic Mode Functions (IMFs) and o residue
  - Deep Belief Network (DBN): includes two Restricted Boltzmann Machines (RBMs) and one single-hidden layer feedforward neural network (SLFN)
  - Linear Neural Network: combines the outputs from DBNs to formulate an ensemble output



1. X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," in Applied Soft Computing, vol. 54, pp. 246–255, 2017. (Codes available)

Experiment Setup

- Datasets
  - The electricity load demand datasets are from Australian Energy Market Operator (AEMO)
  - From three states: New South Wales, Tasmania and Queensland



- Characteristics of Data
  - Seasonality: daily and weekly
  - Three strongest dependent variables:

$$X_{t-1}, X_{t-336}, X_{t-48}$$

  - Guide for feature selection: ACF

$$r_k = r(X_t, X_{t-k}) = \frac{\sum_{t=k+1}^{n}(X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \quad (3.1)$$

where $\bar{X}$ is the mean value of all X in the given time series, r_k measures the linear correlation of the time series at times t and $t - k$.

Experiment Setup

- Normalization: min-max normalization

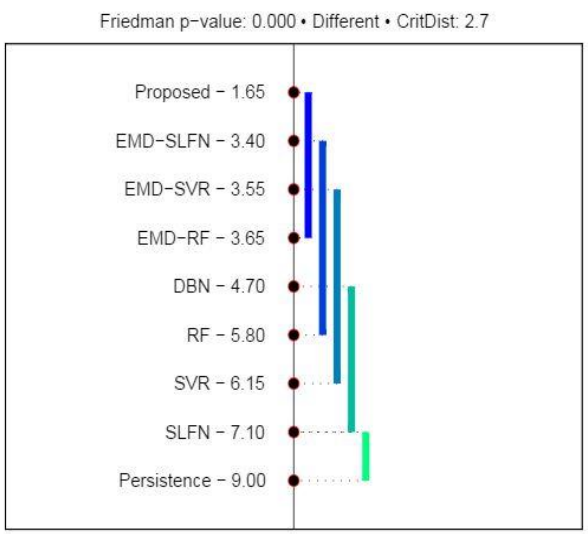
$$\bar{y}_i = \frac{y_{max} - y_i}{y_{max} - y_{min}}$$

y_{max} and y_{min} are the maximum and minimum values of the time series signal. y_i' and y_i are the scaled and original values.

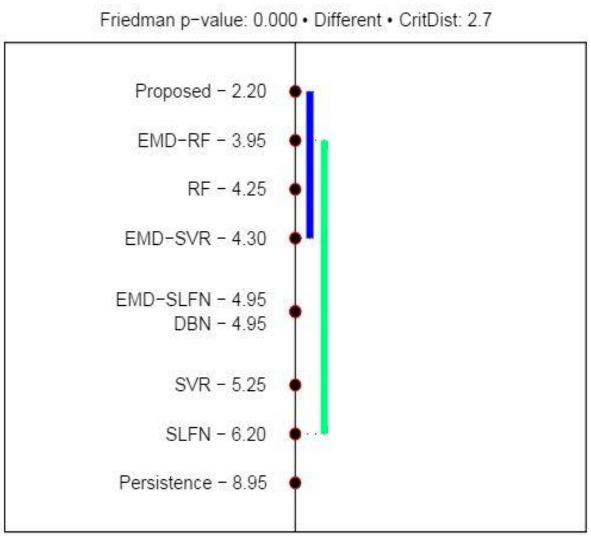
- Implementation
  - SVR and EMD-SVR:
    - LIBSVM
    - RBF kernel
    - Grid search for parameters
  - ANN, DBN, EMD-ANN and Proposed method:
    - Deep learning toolbox in Matlab
    - Number of iterations for BP is searched within [100, 500]
- Performance Estimation: RMSE and MAPE

Statistical Testing for Load Demand Forecasting Results

- Nemenyi testing is applied for statistical testing
- The critical distance is 2.7



Nemenyi testing for half-an-hour ahead load forecasting based on RMSE



Nemenyi testing for one-day ahead load forecasting based on RMSE

- The proposed method has the best rank

NANYANG TECHNOLOGICAL UNIVERSITY

Comparative Experiment I

- Monthly electric load demand data of Northeast China
- Two benchmark methods:
  - Seasonal recurrent SVR with chaotic artificial bee colony (SRSVRCABC), developed by Hong *et al.*, published in *Energy* in 2011
  - Trend fixed seasonal adjustment ε-SVR (TF-ε-SVR-SA), developed by Wang *et al.*, published in *Energy Policy* in 2009

Time point	Actual	ARIMA	TF-ε-SVR-SA	SRSVRCABC	Proposed
October 2008	181.07	192.9316	184.5035	178.4199	181.1451
November 2008	180.56	191.127	190.3608	188.3091	182.4483
December 2008	189.03	189.9155	202.9795	195.3528	184.2608
January 2009	182.07	191.9947	195.7532	187.0825	184.1379
February 2009	167.35	189.9398	167.5795	166.1220	178.7636
March 2009	189.30	183.9876	185.9358	185.1950	184.8475
April 2009	175.84	189.3480	180.1648	179.5335	175.7244
MAPE(%)		6.044	3.799	2.387	**1.998**

Comparative Experiment II

- Electricity load demand data of May 2007 from New South Wales, Australia
- Benchmark method:
  - The adaptive fuzzy combination model (AFCM), developed by Che *et al.*, published in *Energy* in 2012
- Two different sample sizes:
  - one week and three weeks

Sample Size	Metric	SVR	AFCM	Proposed
Small	MAPE	1.3678%	0.9905%	**0.6695%**
	RMSE	145.865	125.323	**83.570**
Large	MAPE	1.6580%	1.2325%	**0.9187%**
	RMSE	181.617	158.754	**118.492**

NANYANG TECHNOLOGICAL UNIVERSITY

Time Series Forecasting

- Introduction
- Data Preprocessing
- Time Series Forecasting with Oblique Random Forest
- RVFL for Forecasting
- EMD based Random Vector Functional Link network & other variants
- Ensemble Incremental Learning with Random Vector Functional Link Network
- Ensemble Deep Learning for Time Series Forecasting
- Empirical Mode Decomposition based Ensemble Deep Learning
- Incremental RVFL based Crude Oil Price Forecasting

Crude Oil Price Forecasting

- Motivation:
  - Crude oil is the most important energy source
  - Relationship between crude oil price and global economic is complicated and nonlinear
  - Some researchers think high oil price has negative relationship with macroeconomic situation, while some others think high oil price is caused by economic growth
- Benchmark models:
  - ANN, SVR
  - CNN, DBN, LSTM
  - We focus on fast non-iterative model

Proposed Incremental RVFL based Ensemble Method

- Off-line Training Procedures:
  - Apply CEEMDAN to decompose crude oil price TS signal into several IMFs and one residue
  - Train an RVFL network for each extracted IMF and residue
  - Aggregate all the outputs by another RVFL to generate the final prediction results.
- Online Updating Procedures:
  - When new input sample occurs, CEEMDAN is re-employed on the signal to get the new input pattern, which can be used to updated the weights of RVFL models.
  - Evaluate the updated model by validation data to determine keep the update or not.

Experiment Setup

- Datasets:
  - Ten years data: 22/01/2008~22/01/2018
  - West Texas Intermediate (WTI)
  - Brent oil
- Forecasting horizon: one day, two days, three days and one week
- Min-max scaling
- Measurements:
  - RMSE
  - MAPE

Oil Price Forecasting Results

TABLE I: Prediction results for crude oil price forecasting

Dataset	Horizon	Metrics	Prediction model							
			Persistence	RVFL [10], [11]	IRVFL [22]	SVR [23]	LSTM [6]	EMD-RVFL [24]	EMD-LSTM [25]	EMD-IRVFL (Proposed)
WTI	1 day	RMSE	2.099	1.114	1.062	0.958	0.753	0.954	**0.747**	0.798
		MAPE	3.523%	1.910%	1.793%	1.628%	1.324%	1.600%	**1.321%**	1.341%
	2 days	RMSE	2.497	1.475	1.475	1.369	**1.001**	1.356	1.003	1.015
		MAPE	4.242%	2.542%	2.530%	2.352%	1.692%	2.330%	1.701%	**1.686%**
	3 days	RMSE	2.764	1.745	1.741	1.645	1.157	1.631	**1.149**	1.161
		MAPE	4.771%	3.071%	3.036%	2.859%	1.832%	2.851%	**1.821%**	1.950%
	1 week	RMSE	3.772	2.600	2.529	3.237	1.512	2.463	**1.483**	1.495
		MAPE	6.889%	4.645%	4.480%	5.403%	2.582%	4.367%	**2.479%**	2.481%
Brent	1 day	RMSE	1.277	1.202	1.164	1.047	**0.812**	1.014	0.822	0.832
		MAPE	2.155%	1.973%	1.890%	1.701%	**1.391%**	1.610%	1.392%	1.421%
	2 days	RMSE	1.872	1.587	1.550	1.395	1.201	1.409	**1.128**	1.139
		MAPE	2.791%	2.608%	2.506%	2.256%	1.853%	2.303%	**1.694%**	1.713%
	3 days	RMSE	2.347	1.915	1.894	1.704	1.408	1.733	**1.387**	1.398
		MAPE	3.564%	3.167%	3.072%	2.765%	**1.982%**	2.872%	2.023%	2.012%
	1 week	RMSE	3.120	2.867	2.718	2.446	**1.602**	2.611	1.682	1.673
		MAPE	4.981%	4.751%	4.464%	4.017%	2.657%	4.322%	2.688%	**2.631%**

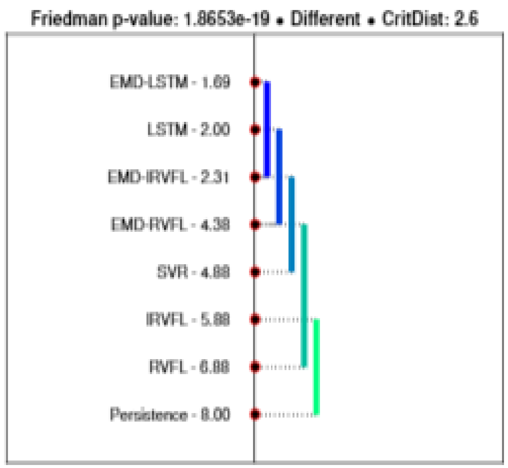


Fig. 2: Nemenyi test for crude oil price forecasting. The critical distance is 2.6.

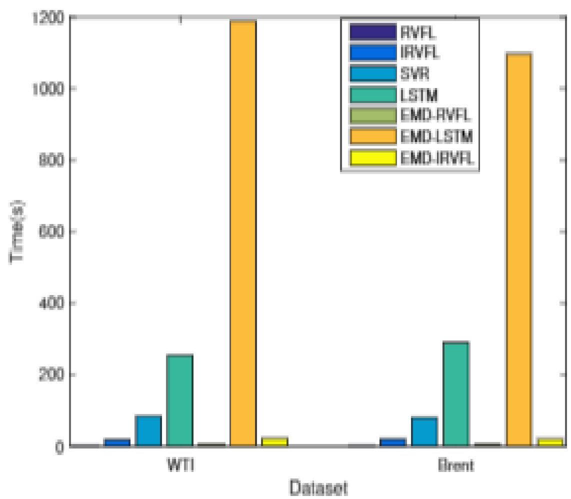


Fig. 3: Computation time of learning models for crude oil price forecasting

Summary:

- Proposed ensemble model outperforms single structure models significantly
- Proposed model cannot beat LSTM and EMD-LSTM based on forecasting accuracy
- The proposed ensemble model has its advantage of fast computing compared with LSTM and EMD-LSTM

NANYANG TECHNOLOGICAL UNIVERSITY

Thank You !

Questions?