Multilevel Wavelet Decomposition Network for Interpretable Time Series Analysis

Jingyuan Wang, Ze Wang, **Jianfeng Li**, Junjie Wu

Beihang University

July 22, 2018

Outline

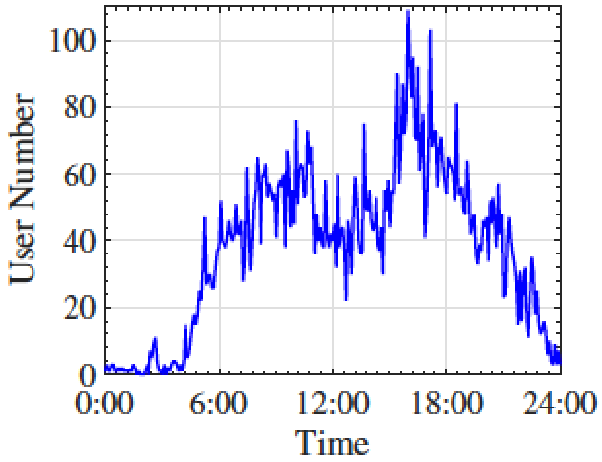
- Part I： Time series analysis and model inspiration
- Part II： mWDN framework with two model structures for different problems
- Part III： Head for interpretability-Importance analysis

Time Series Analysis

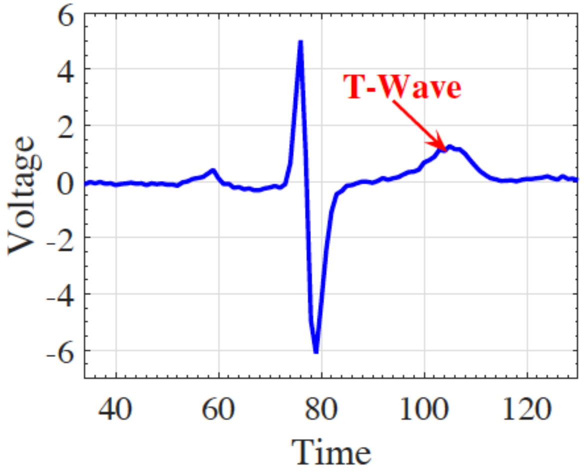
- Time Series

$$X = \{X_1, X_2, \ldots\}.$$

- Time Series Analysis
  - Classification
  - Forecasting
  - Abnormal detection
  - Query by content
  - …



Cell-phone User Number

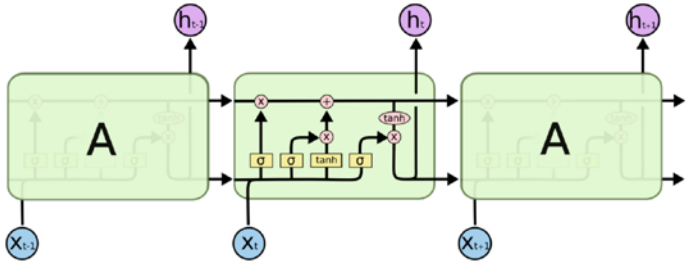


ECG Data

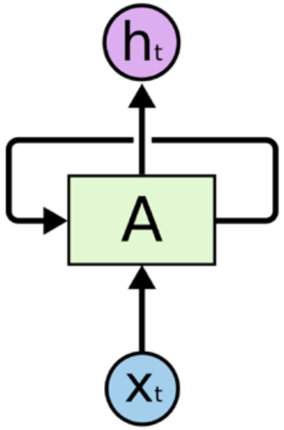
....

Deep Models for Time Series Analysis

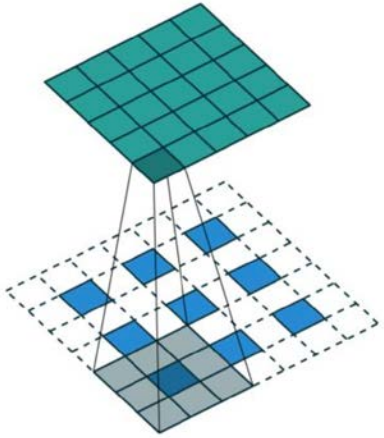
Time Domain



LSTM



RNN

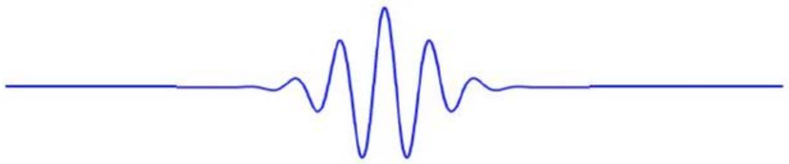


CNN

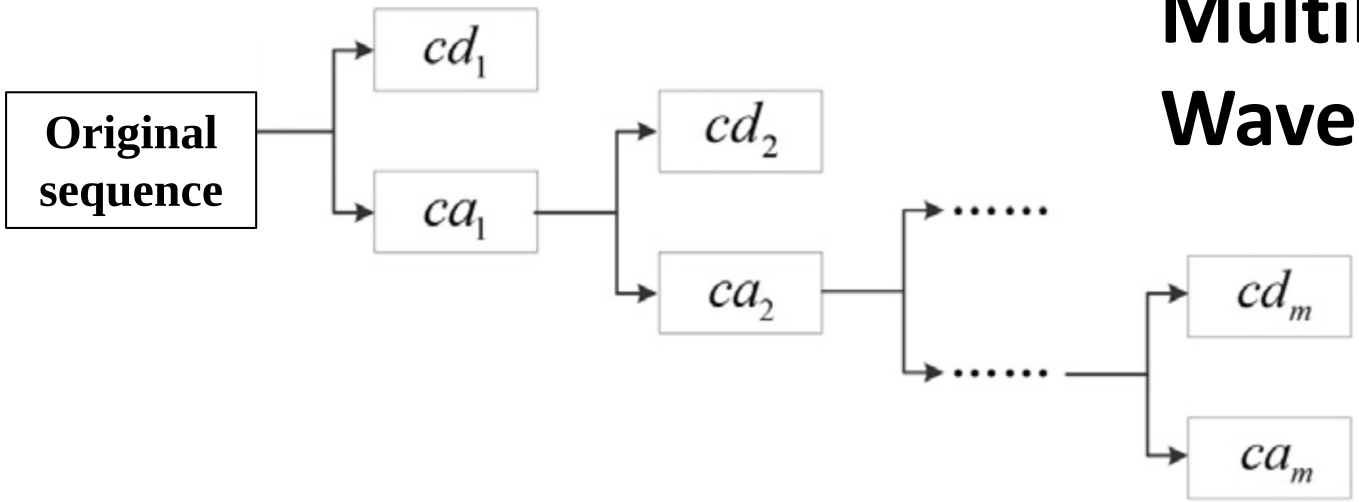
Frequency Domain

?

Time Series Analysis Best Practice



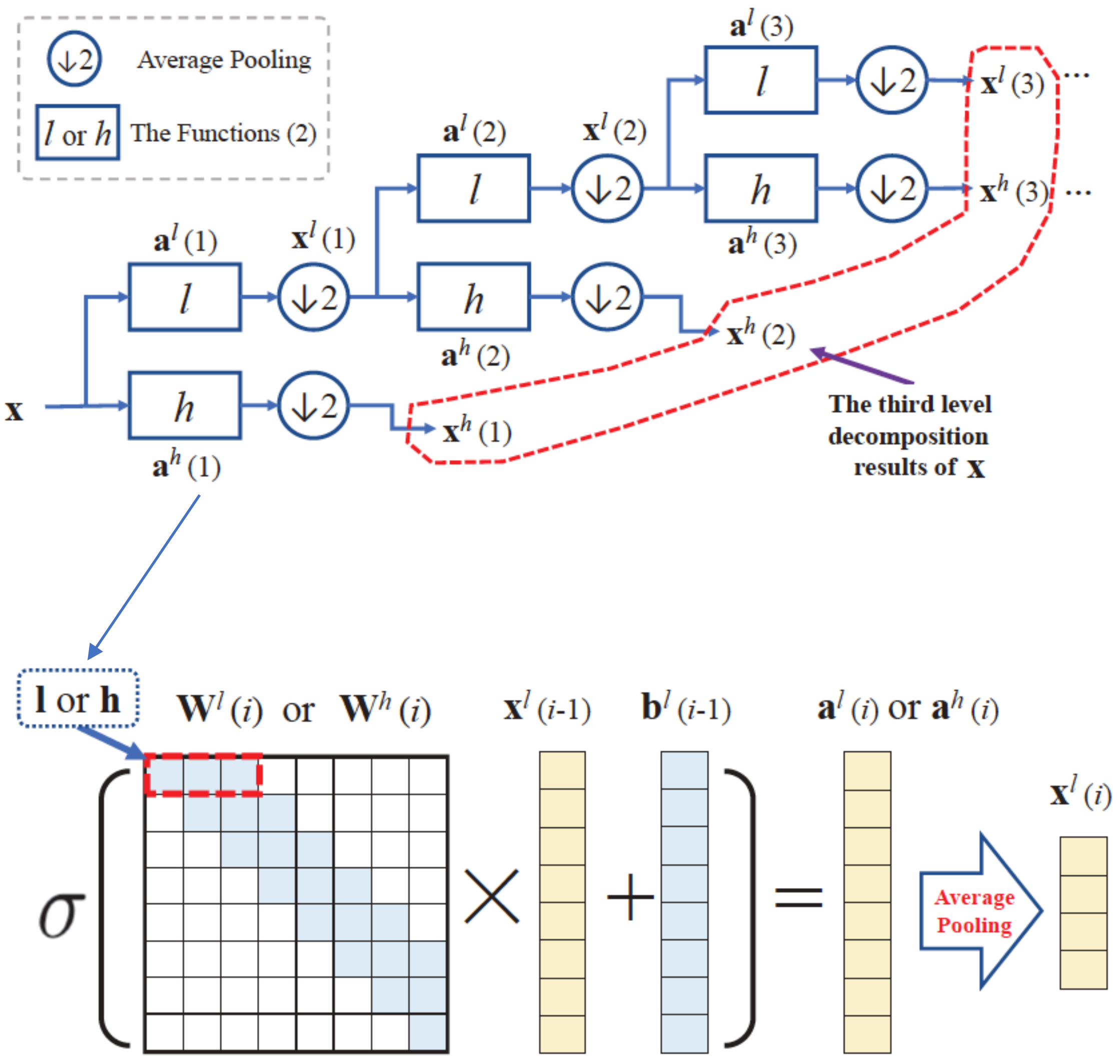
Wavelet Decomposition



Multilevel Wavelet Decomposition

We propose a **wavelet-based** neural network structure called multilevel Wavelet Decomposition Network (mWDN) for building **frequency-aware** deep learning models for time series analysis.

Multilevel Wavelet Decomposition Network (mWDN)



- Difference:
  - Weight parameters in mWDN are initialized with db4 wavelet basis.
  - All other epsilon weight parameters initialized with random small value.
  - All weight parameters in mWDN can be fine-tuned,.
  - Build a seamless end-to-end learning framework.

$$\mathbf{W}^l(i) = \begin{bmatrix} l_1 & l_2 & l_3 & \cdots & l_K & \epsilon & \cdots & \epsilon \\ \epsilon & l_1 & l_2 & \cdots & l_{K-1} & l_K & \cdots & \epsilon \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & l_1 & \cdots & l_{K-1} & l_K \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & \cdots & \cdots & l_1 & l_2 \\ \epsilon & \epsilon & \epsilon & \cdots & \cdots & \cdots & \epsilon & l_1 \end{bmatrix},$$

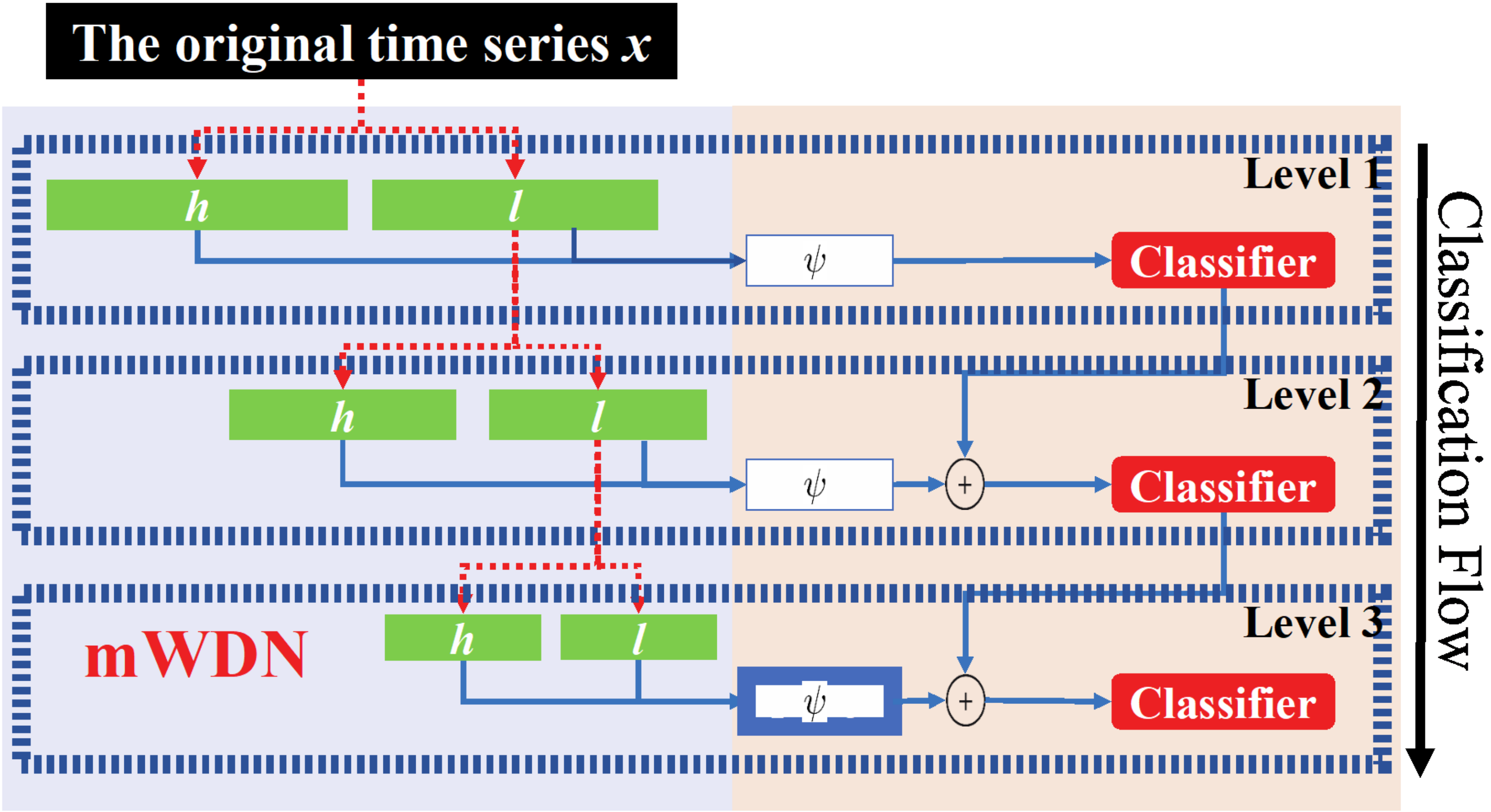
$$\mathbf{a}^l(i) = \sigma\left(\mathbf{W}^l(i)\mathbf{x}^l(i-1) + \mathbf{b}^l(i)\right),$$
$$\mathbf{a}^h(i) = \sigma\left(\mathbf{W}^h(i)\mathbf{x}^l(i-1) + \mathbf{b}^h(i)\right),$$

$$\mathbf{W}^h(i) = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_K & \epsilon & \cdots & \epsilon \\ \epsilon & h_1 & h_2 & \cdots & h_{K-1} & h_K & \cdots & \epsilon \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & h_1 & \cdots & h_{K-1} & h_K \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & \cdots & \cdots & h_1 & h_2 \\ \epsilon & \epsilon & \epsilon & \cdots & \cdots & \cdots & \epsilon & h_1 \end{bmatrix}.$$

Part II：mWDN Framework With Two Model Structures for Different Problems

- multi-level Wavelet Decomposition Network（mWDN）
- For classification：Residual Classification Flow network（RCF）
- For forecasting：multi-frequency Long Short-Term Memory network（mLSTM）

Residual Classification Flow Network (RCF)



- Forward Network

$$\mathbf{u}(i) = \psi\left(\mathbf{x}^h(i), \mathbf{x}^l(i), \theta^\psi\right),$$

- Residual Learning Manner

$$\hat{\mathbf{c}}(i) = S\left(\hat{\mathbf{c}}(i-1) + \mathbf{u}(i)\right),$$

Classification Problem: UCR Dataset

Name	First paper or data creator	Number of classes	Size of training set	Size of testing set	Time series Length
Synthetic Control	Pham	6	300	300	60
Gun–Point	Ratanamahatana	2	50	150	150
CBF		3	30	900	128
Face (all)	Xi	14	560	1690	131
OSU Leaf	Gandhi	6	200	242	427
Swedish Leaf	Soderkvist	15	500	625	128
50Words	Rath	50	450	455	270
Trace	Roverso	4	100	100	275
Two Patterns	Geurts	4	1000	4000	128
Wafer	Olszewski	2	1000	6174	152
Face (four)	Ratanamahatana	4	24	88	350
Lightning–2	Eads	2	60	61	637
Lightning–7	Eads	7	70	73	319
ECG	Olszewski	2	100	100	96
Adiac	Jalba	37	390	391	176
Yoga	Xi	2	300	3000	426
Fish	Lee	7	175	175	463
Plane		7	105	105	144
Car		4	60	60	577
Beef	Tony Bagnall	5	30	30	470

- Test on 40 UCR datasets.
- Competitor models.
  - RNN, LSTM
  - MLP, FCN, ResNet
  - MLP-RCF, FCN-RCF, ResNet-RCF
  - Wavelet-RCF
- 10 times for each model.
- Evaluation (MPCE):

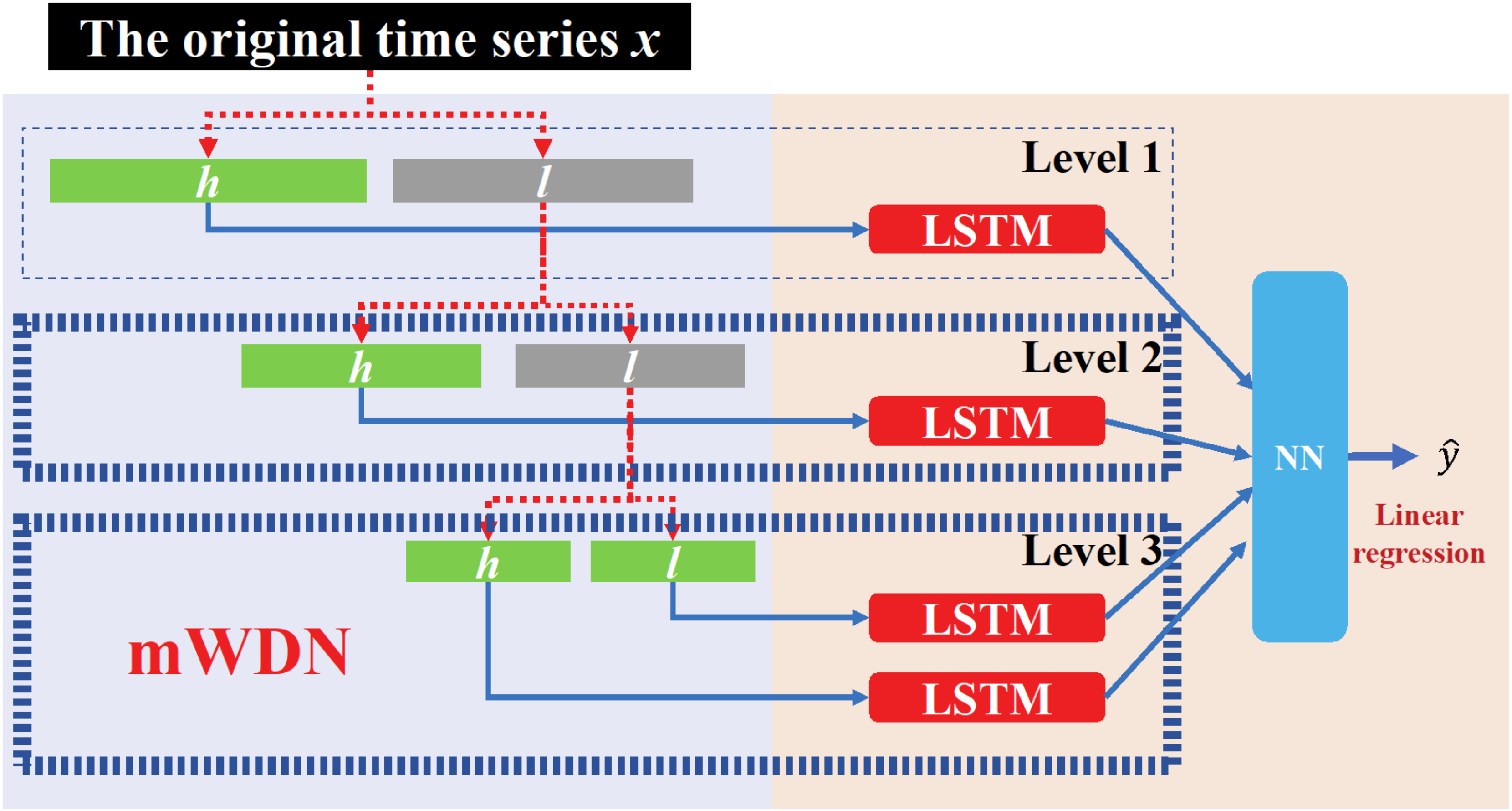
$$\mathrm{MPCE} = \frac{1}{K}\sum_{l=1}^{K}\frac{e_k}{C_k}.$$

Err Rate	RNN	LSTM	MLP	FCN	ResNet	**MLP-RCF**	**FCN-RCF**	**ResNet-RCF**	Wavelet-RCF
Adiac	0.233	0.341	0.248	**0.143**	0.174	0.212	0.155	*0.151*	0.162
Beef	0.233	0.333	0.167	0.25	0.233	*0.06*	**0.03**	*0.06*	*0.06*
CBF	0.189	0.118	0.14	**0**	*0.006*	0.056	**0**	**0**	0.016
ChlorineConcentration	0.135	0.16	0.128	0.157	0.172	0.096	**0.068**	*0.07*	0.147
CinCECGtorso	0.333	0.092	0.158	0.187	0.229	0.117	*0.014*	0.084	**0.011**
CricketX	0.449	0.382	0.431	*0.185*	**0.179**	0.321	0.216	0.297	0.211
CricketY	0.415	0.318	0.405	0.208	0.195	0.254	**0.172**	0.301	*0.192*
CricketZ	0.4	0.328	0.408	*0.187*	0.187	0.313	**0.162**	0.275	**0.162**
DiatomSizeReduction	0.056	0.101	0.036	0.07	0.069	**0.013**	*0.023*	0.026	0.028
ECGFiveDays	0.088	0.417	0.03	*0.015*	0.045	0.023	**0.01**	0.035	0.016
FaceAll	0.247	0.192	0.115	**0.071**	0.166	0.094	0.098	0.126	*0.076*
FaceFour	0.102	0.364	0.17	0.068	0.068	0.102	**0.05**	*0.057*	0.058
FacesUCR	0.204	0.091	0.185	*0.052*	**0.042**	0.15	0.087	0.102	0.087
50words	0.316	0.284	0.288	0.321	*0.273*	0.316	0.288	**0.258**	0.3
FISH	0.126	0.103	0.126	0.029	**0.011**	0.086	*0.021*	0.034	0.026
GunPoint	0.1	0.147	0.067	**0**	*0.007*	0.033	**0**	0.02	**0**
Haptics	0.594	0.529	0.539	**0.449**	0.495	0.480	*0.461*	0.473	0.476
InlineSkate	0.667	0.638	0.649	0.589	0.635	**0.543**	*0.566*	0.578	0.572
ItalyPowerDemand	0.055	0.072	0.034	0.03	0.04	0.031	**0.023**	0.034	*0.028*
Lighting2	**0**	**0**	0.279	0.197	0.246	0.213	*0.145*	0.197	0.162
Lighting7	0.288	0.384	0.356	*0.137*	0.164	0.179	**0.091**	0.177	0.144
MALLAT	0.119	0.127	0.064	**0.02**	*0.021*	0.058	0.044	0.046	0.024
MedicalImages	0.299	0.276	0.271	0.208	0.228	0.251	**0.164**	*0.188*	0.206
MoteStrain	0.133	0.167	0.131	*0.05*	0.105	0.105	0.076	**0.032**	0.05
NonInvasiveFatalECGThorax1	0.09	0.08	0.058	0.039	0.052	*0.029*	**0.026**	0.04	0.042
NonInvasiveFatalECGThorax2	0.069	0.071	0.057	0.045	0.049	0.056	**0.028**	*0.033*	0.048
OliveOil	0.233	0.267	0.6	0.167	0.133	0.03	**0**	**0**	*0.012*
OSULeaf	0.463	0.401	0.43	**0.012**	0.021	0.342	*0.018*	0.021	0.021
SonyAIBORobotSurface	0.21	0.309	0.273	*0.032*	**0.015**	0.193	0.042	0.032	0.052
SonyAIBORobotSurfaceII	0.219	0.187	0.161	**0.038**	**0.038**	0.092	*0.064*	0.083	0.072
StarLightCurves	0.027	0.035	0.043	0.033	0.029	*0.021*	**0.018**	0.027	0.03
SwedishLeaf	0.085	0.128	0.107	*0.034*	0.042	0.089	0.057	**0.017**	0.046
Symbols	0.179	0.117	0.147	**0.038**	0.128	0.126	*0.04*	0.107	0.084
TwoPatterns	0.005	*0.001*	0.114	0.103	**0**	0.070	**0**	**0**	0.005
uWaveGestureLibraryX	0.224	0.195	0.232	0.246	0.213	0.213	0.218	*0.194*	**0.162**
uWaveGestureLibraryY	0.335	0.265	0.297	0.275	0.332	0.306	**0.232**	0.296	*0.241*
uWaveGestureLibraryZ	0.297	0.259	0.295	0.271	0.245	0.298	0.265	*0.204*	**0.194**
wafer	**0**	**0**	0.004	*0.003*	*0.003*	*0.003*	**0**	**0**	**0**
WordsSynonyms	0.429	0.343	0.406	0.42	0.368	0.391	*0.338*	0.387	**0.314**
yoga	0.202	0.158	0.145	0.155	0.142	0.138	**0.112**	0.139	*0.128*
Winning times	2	2	0	*9*	6	2	**19**	7	7
AVG arithmetic ranking	7.425	6.825	7.2	4.025	4.55	5.15	**2.175**	3.375	*3.075*
AVG geometric ranking	6.860	6.131	7.043	3.101	3.818	4.675	**1.789**	2.868	*2.688*
MPCE	0.039	0.043	0.041	0.023	0.025	0.028	**0.017**	0.021	*0.019*

Part II：mWDN Framework With Two Model Structures for Different Problems

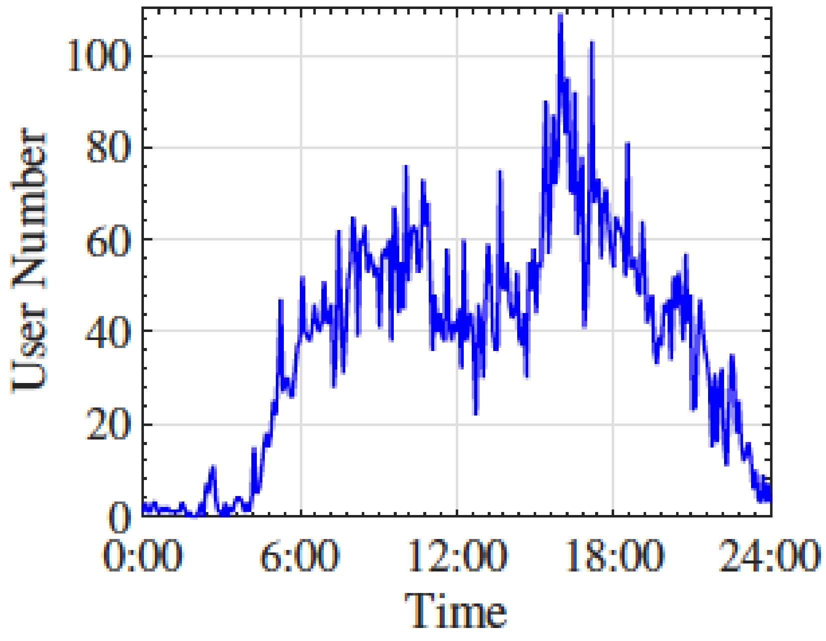
- multi-level Wavelet Decomposition Network（mWDN）
- For classification：Residual Classification Flow network（RCF）
- For forecasting：multi-frequency Long Short-Term Memory network（mLSTM）

Multi-frequency Long Short-Term Memory network (mLSTM)



- Using N+1 LSTM to forecasts the future state of one sub-series in X(N).
- Fuse the LSTM output with a fully connected neural network.

Forecasting Problem: Wuxi Cell-phone User Data



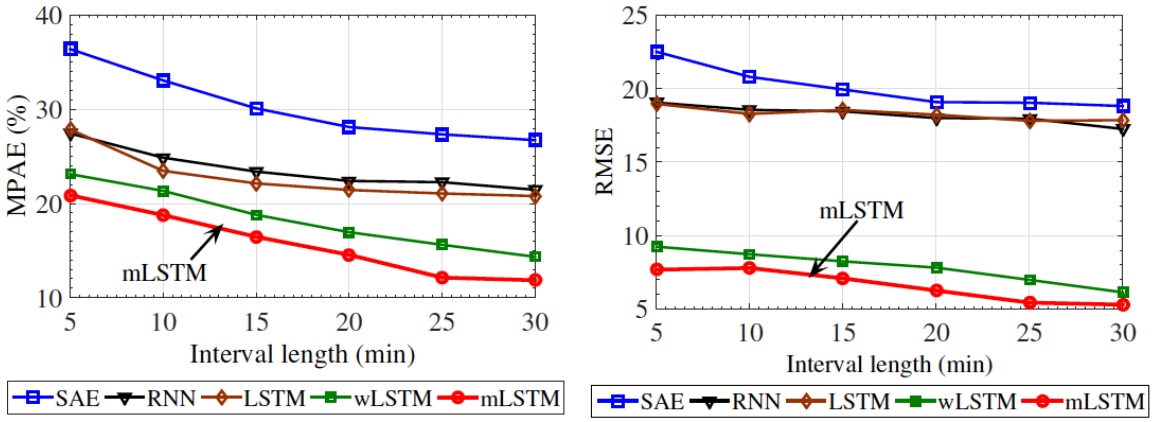
Cell-phone User Number

- Test on real-life *WuxiCellPhone* datasets.
  - From 20 base stations during 2 weeks.
  - Time granularity is 5 minutes.
- Competitor models.
  - SAE
  - RNN, LSTM
  - wLSTM
- Evaluation.
  - MAPE
  - RMSE

$$\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\frac{|\hat{x}_t - x_t|}{x_t} \times 100\%,$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{x}_t - x_t)^2},$$

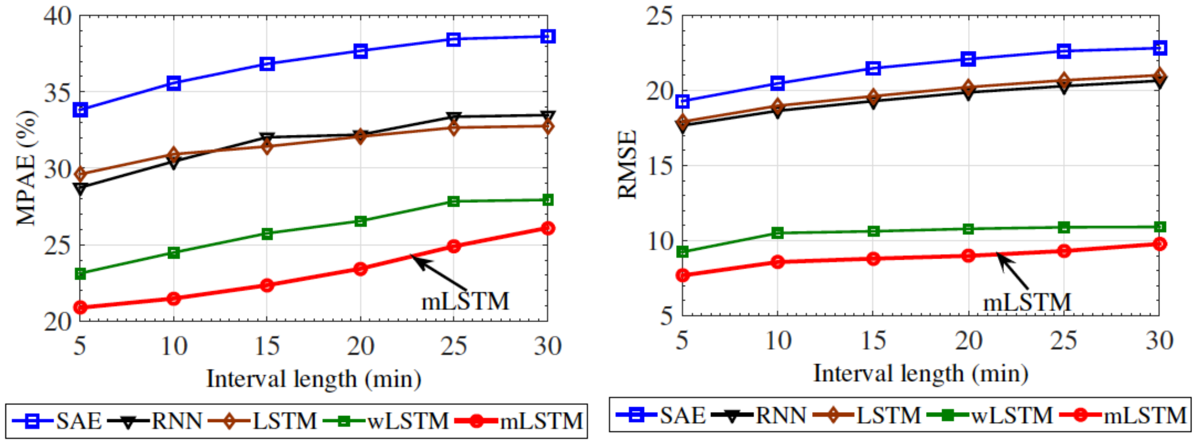
Forecasting Problem: Cell-phone User Number Forecasting in Wuxi



(a) Comparison by MAPE (b) Comparison by RMSE

Varying period lengths

Varying interval lengths



(a) Comparison by MAPE (b) Comparison by RMSE

Outline

- Part I： Time series analysis and model inspiration
- Part II： mWDN framework with two model structures for different problems
- Part III： Head for interpretability-Importance analysis

Importance Analysis

- **Define importance of inputs as well as sub-series nodes:**

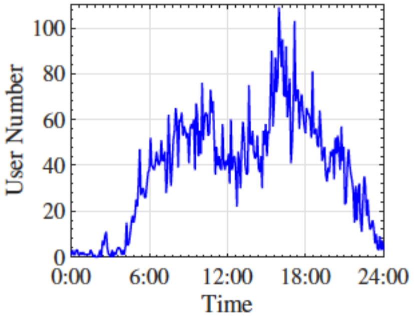
$$S_a(\mathbf{x}) = \left|\frac{\partial M(a(\mathbf{x}))}{\partial a(\mathbf{x})}\right| = \left|\lim_{\varepsilon \to 0} \frac{M(a(\mathbf{x})) - M(a(\mathbf{x}) - \varepsilon)}{\varepsilon}\right|.$$

- **Importance of a node with respect to a training dataset:**

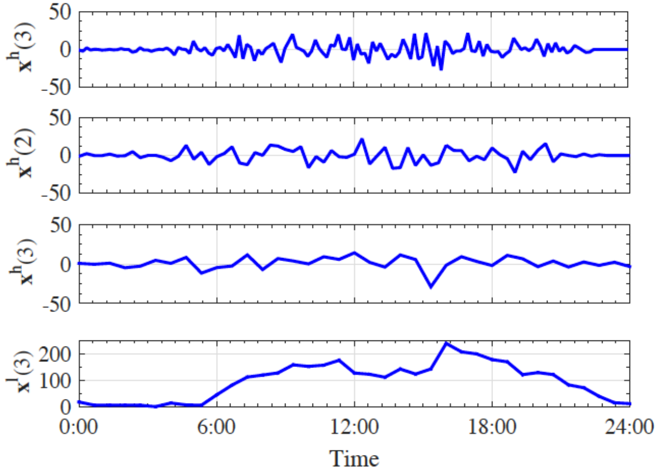
$$I(a) = \frac{1}{J} \sum_{j=1}^{J} S_a(\tilde{\mathbf{x}}^j).$$

We **quantify the importance** of each middle layer to the final output of the mWDN based models by estimating the **partial derivatives** of final prediction to intermediate sub-sequence generated by Multilevel Wavelet Decomposition Network.

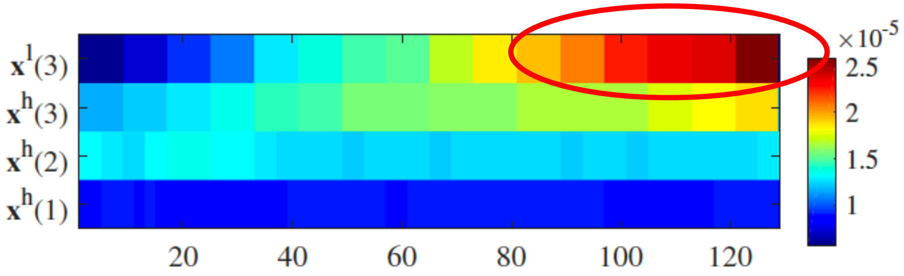
Forecasting: Cell-phone User Number in Wuxi with Layer Importance Analysis



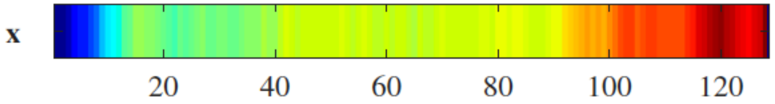
Cell-phone User Number



Cell-phone User Numbers in Different Layers

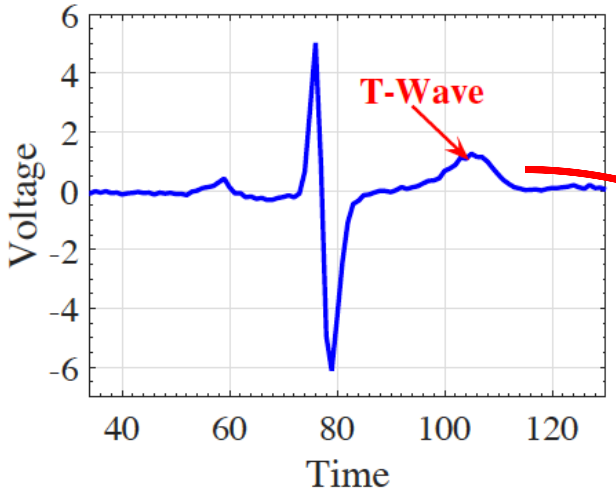


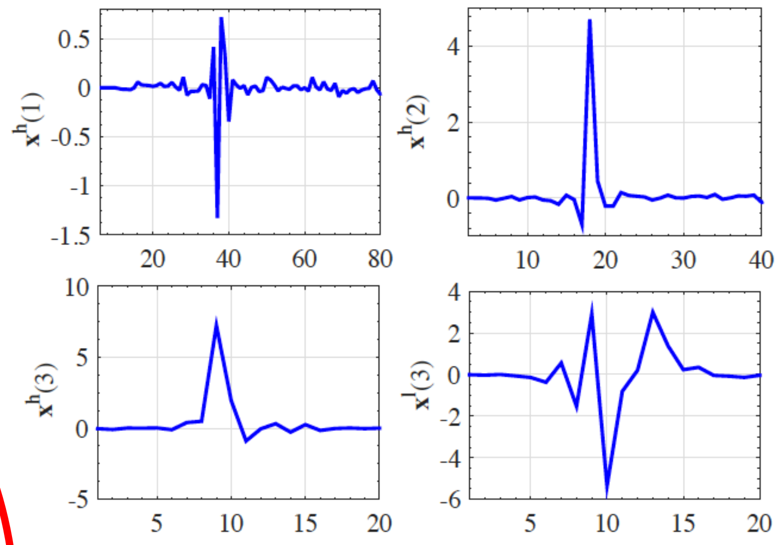
(a) Importance spectra of middle layers



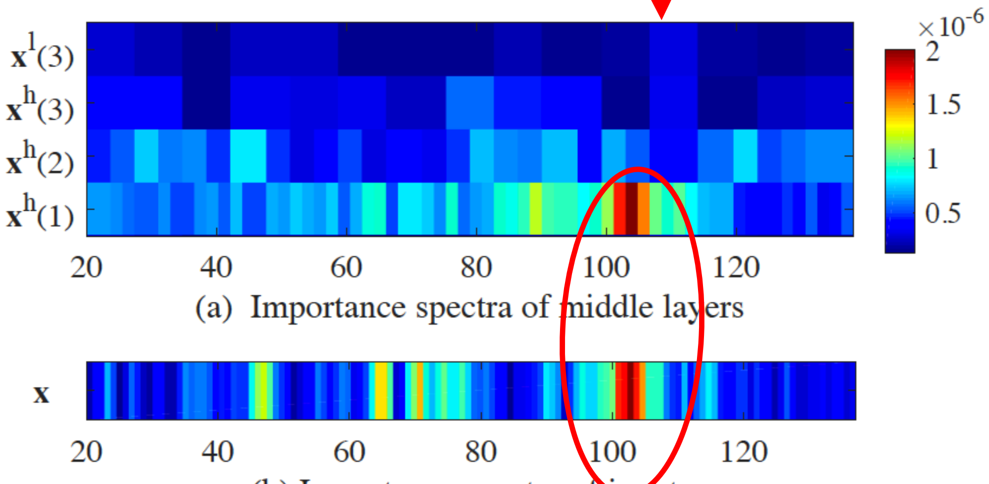
(b) Importance spectra of inputs

Classification: ECG Data of UCR with Layer Importance Analysis





(b) ECG Waves in Different Layers



(a) Importance spectra of middle layers

(b) Importance spectra of inputs

Q&A

Thanks!