TATA CONSULTANCY SERVICES
Experience certainty.

ICASSP 2020

Paper# 3324

Registration# 2036

Instant Adaptive Learning: An Adaptive Filter Based Fast Learning Model Construction for Sensor Signal Time Series Classification on Edge Devices

Arpan Pal[1], **Arijit Ukil[1]**, Trisrota Deb[1], Ishan Sahu[1], Angshul Majumdar[2]

[1]Research and Innovation, Tata Consultancy Services, Kolkata, India
[2]Indraprastha Institute of Information Technology, Delhi, India

Presenter- Arijit Ukil

Summary of the work

- **Construction of learning model under computational and energy constraints for practical IoT time series sensor signal analytics applications over edge devices**

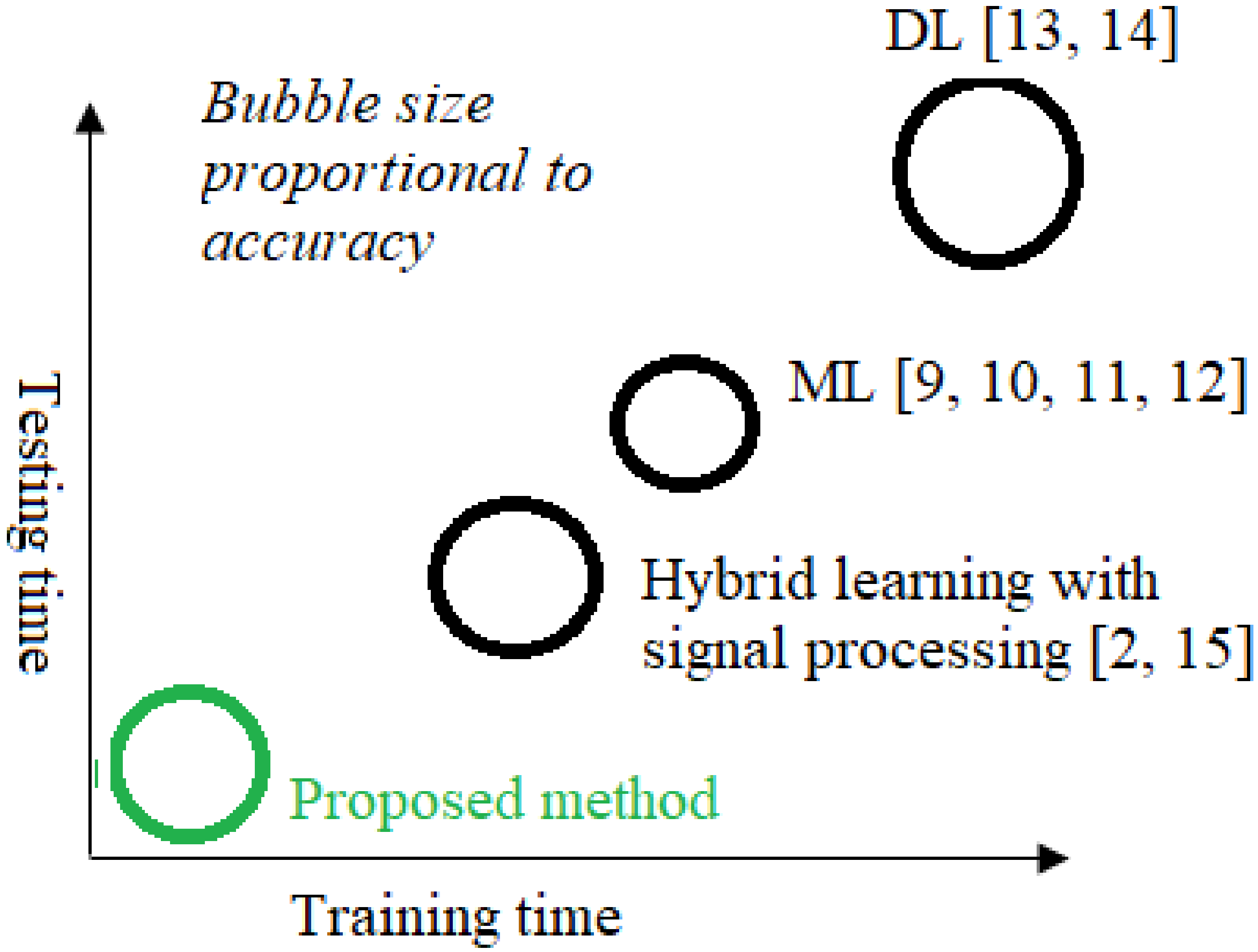
- **Majority of the state-of-the-art algorithms and solutions attempt to achieve high performance objective (like test accuracy) irrespective of the computational constraints of real-life applications**

- **We propose Instant Adaptive Learning that characterizes**
  - ✓ **intrinsic signal processing properties of time series sensor signals using linear adaptive filtering**
  - ✓ **derivative spectrum to efficiently construct the feature space for low-cost learning model followed by standard classification algorithms**

Prior works

- Dynamic time warping based distance measures with nearest neighbor classifier- DTW-1NN [9]: Performance is poor, DTW measure is computationally expensive
- Time series substructure is learnt using symbolic Fourier approximation- *BOSS* [10]: Performance is moderate, computationally expensive
- Large number of classifiers with each being hyperparameter optimized: *HIVE-COTE* and *COTE* [11]: Performance is good, computationally hugely expensive
- Deep neural network approach like Residual Network (*ResNet*) [13- 14]: Performance is good, computationally hugely expensive, GPU is required for training
- Signal Processing based Generic Feature (SPGF) with TimeNet (TN) [2, 15]: a hybrid approach of fusing signal processing based features and pre-trained features: Performance is good, computational requirement is moderate

The edge learning landscape



ICASSP 2020

Representative challenges

TCS RESEARCH & INNOVATION

- Complete on-board/ on-device processing with constrained computational resource (Memory, Compute time, Energy)
- Diverse Scenarios – significant time spent in signal conditioning and feature engineering
- Near-real Time Response – Instant inference
- Online near real-time Learning – Instant training

Our contribution

TCS RESEARCH & INNOVATION

We propose a novel adaptive filter based approach: Instant Adaptive Learning, which enables automated and computationally lightweight learning (significant reduction in computational resources) while marginally trading off its performance

Our contribution

- It automates the feature engineering via adaptive filter construction followed by simple spectral features derived from the adaptive filter output and error signals
- It makes both learning and inferencing to be extremely lightweight from compute, memory and power perspectives
- It has provision to take metadata or expert knowledge involving frequency response of the given system as input to impact the learnability of the model

Novelty of this work

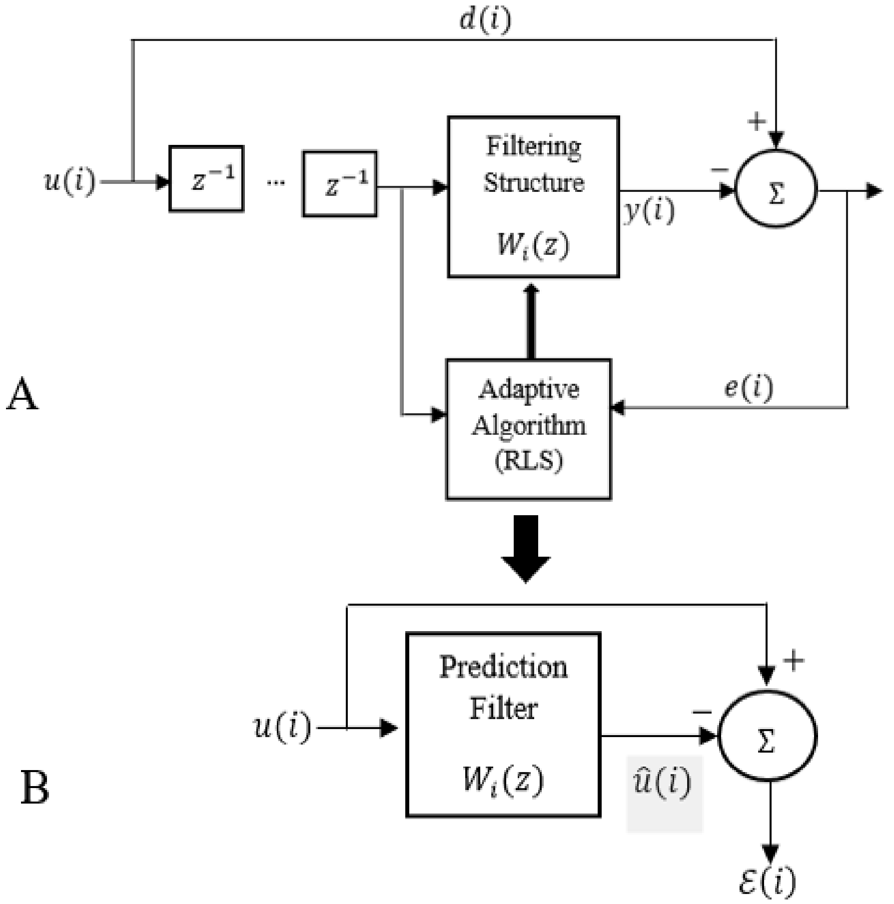
TCS RESEARCH & INNOVATION

- The main novelty of this paper is the representation capability of adaptive filtering based learning for time series classification under highly constrained computational budget
- The proposed Instant Adaptive Learning method is a novel signal processing approach that attempts to characterize the training signal distribution under the hypothesis
  - ✓ an adaptive filter is a reliable estimator of input time series signal, when for the desired signal, a time advanced version of the input signal is given

Methodology- Prediction Filter Construction

TCS RESEARCH & INNOVATION

- d(i) is the desired signal for constructing the RLS filter
- u(i) is the input (training) signal
  - ✓ u(i) is a time delayed version of d(i)
- For each of the training signal, the prediction filter is constructed – Figure A
- Subsequently, each of the training signals u(i) is passed to its prediction filter – Figure B
  - ✓ The output of the prediction filter:
    - ➢ $\hat{u}(i): u(i) \xrightarrow{W_i(z)} \hat{u}(i)$
    - ➢ $\mathcal{E}(i) = u(i) - \hat{u}(i)$

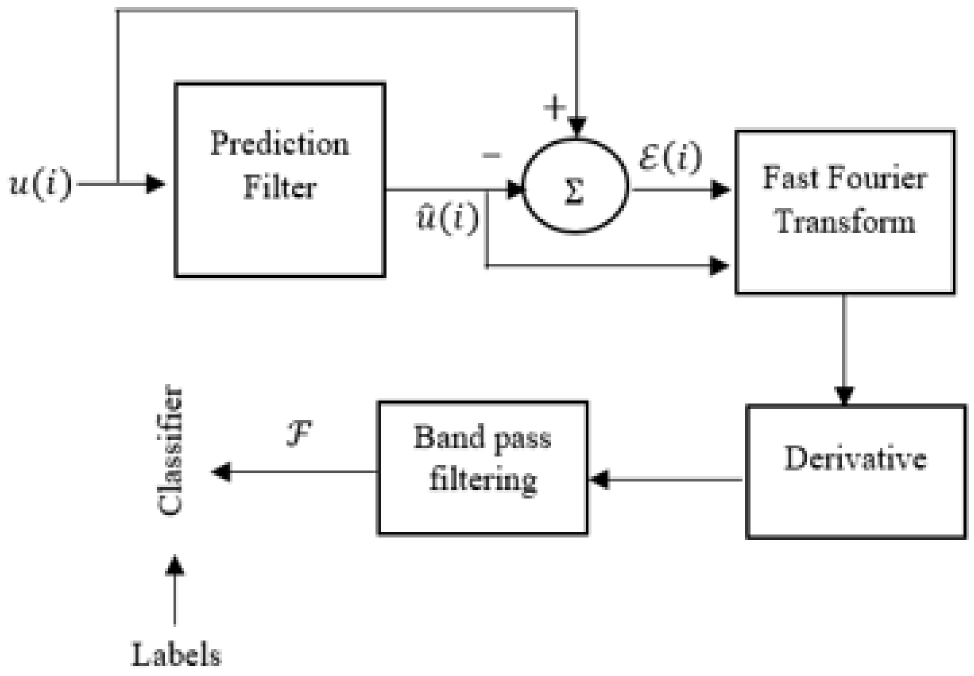


Methodology- Prediction Filter Construction

TCS RESEARCH & INNOVATION

- According to the Wold decomposition [17] : $u(i) = u_d(i) + u_s(i)$, where, $u_d(i)$ denotes the deterministic or regular part of $u(i)$, and $u_s(i)$ is the stochastic or random part of $u(i)$.
  - ✓ In our case, $\hat{u}(i)$ represents the deterministic component ($u_d(i)$) and $\mathcal{E}(i)$ is the stochastic component ($u_s(i)$)
- **Feature selection-** we use a widely used spectrum characterization technique called derivative spectrum - the differentiation of spectral coefficients [18]
  - ✓ $\{\hat{u}, \mathcal{E}\} \xrightarrow{\substack{Derivative \\ spectrum}} \{\hat{\mathbb{U}}, \mathbb{E}\}$
  - ✓ $\{\hat{u}, \mathcal{E}\} \xrightarrow{\substack{Band \\ limited\ by\ expert \\ driven\ cut-off\ (\Omega_l, \Omega_u)}} \{\mathcal{F}_{\hat{\mathbb{U}}}, \mathcal{F}_{\mathbb{E}}\}$

✓ We use standard classification algorithm like SVM with Radial Basis Function [19] as the kernel with grid-searched hyperparameters to generate the training model using the feature vector $\mathcal{F} = \{\mathcal{F}_{\hat{\mathbb{U}}}, \mathcal{F}_{\mathbb{E}}\}$.



ICASSP 2020

Methodology- Testing

- **Feature selection process is unsupervised**
- Each of the test signals are transformed to $\hat{u}_{test}, \mathcal{E}_{test}$ and transformation to $\widehat{\mathcal{U}}_{test}, \mathbb{E}_{test}$ are done
- Test feature vector $\mathcal{F}_{test}$ is generated from $\widehat{\mathcal{U}}_{test}, \mathbb{E}_{test}$
- The generated test features $\mathcal{F}_{test}$ are fed to the constructed trained model to infer the class of the test instances

Results

- We experiment with number of time series sensor signals available from open access database – UCR time series archive [7, 8]
- The five dataset are: FordA, FordB, wafer, Earthquake and ECGFiveDays
- These dataset represent different application scenarios for practical time series analysis
  - FordA and FordB try to identify faulty automobile engines analyzing the engine noise
  - Wafer dataset represents semiconductor process control measurements of normal and abnormal fabrication
  - Earthquake dataset consists of vibration signals correspond to earthquake vents
  - ECGFiveDays is a set of Electrocardiogram (ECG) signals of cardiac activity problem
- The development environment is x86 architecture with 64-bit CPU and 16 cores of Intel Xeon CPU E5-2623 v4 with 2.60GHz clock speed
- The software development is performed on Release 2017b of MATLAB

Results

Adaptive filter hyper-parameters for different datasets

UCR dataset	Filter Order (p)	Cut-off (Ω_l, Ω_u)	Delay (δ)
FordB	80	[0.44, 1]	1
FordA	64	[0.44, 1]	4
wafer	16	[0.39, 1]	4
Earthquake	16	[0.39, 1]	1
ECGFiveDays	64	[0.39, 1]	4

Classification performance in terms of accuracy (in %)

UCR dataset	TN-C [2, 15]	SPGF-TN-C [2]	DTW-1NN [9]	Instant Adaptive Learning (Proposed)
FordB	73.70	89.25	68.60	**80.31**
FordA	78.10	93.17	65.90	64.00
wafer	99.50	99.88	99.50	95.47
Earthquake	76.70	81.98	70.30	**81.98**
ECGFiveDays	92.60	94.89	79.70	75.26

Comparison - total training time (in seconds)

UCR dataset	TN-C [2, 15]	SPGF-TN-C [2]	Instant Adaptive Learning (Proposed)	
			Time (sec)	Speed-up wrt TN-C
FordB	2103	2897	11.96	175.8
FordA	2265	3542	19.96	113.5
wafer	2113	2824	5.19	407.1
Earthquake	1983	2108	2.91	681.4
ECGFiveDays	1984	1998	0.365	5435.6

Comparison - per instance inferencing time (in msec)

UCR dataset	TN-C [2, 15]	SPGF-TN-C [2]	Instant Adaptive Learning (Proposed)	
			Time (msec)	Speed-up wrt TN-C
FordB	379	835	16	23.7
FordA	458	875	16	28.6
wafer	319	657	5	63.8
Earthquake	331	929	1.98	167.2
ECGFiveDays	471	721	6	78.5

Conclusion

TCS RESEARCH & INNOVATION

- Instant Adaptive Learning is poised to satisfy the guarantee of faster learning and inferencing through novel adaptive filtering and derivative spectrum based exploration of sensor signal property estimation
- This approach and the proposed algorithm are appropriate for gamut of IoT applications for edge learning
  - ✓ E-health, earthquake detection, real-time machine inspection at factories with restricted access and hazardous environment
- We intend to conclude **that the proposed simple, efficient learning has the potential to be deployed in sensor nodes, edge devices, even directly on the sensing devices that possess tiny computing power but require faster learning capability**

Thank You

contact: arijit.ukil@tcs.com

Reference

[1] A. Pal, P. Balamuralidhar, "Iot Technical Challenges and Solutions," Artech House, 2016.
[2] A. Ukil, P. Malhotra, S. Bandyopadhyay, T. Bose, I. Sahu, A. Mukherjee, L. Vig, A. Pal, and G. Shroff, "Fusing Features based on Signal Properties and TimeNet for Time Series Classification," 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 2019.
[3] S. Banerjee, T. Chattopadhyay, A. Pal, and U. Garain, "Automation of feature engineering for IoT analytics," ACM SIGBED Review, pp. 24-30, 2018.
[4] S. Haykin, Adaptive Filter Theory, 4th Edition, Upper Saddle River, NJ: Prentice Hall, 2002.
[5] S.C. Douglas, "Numerically-robust O(N^2) RLS algorithms using least-squares prewhitening," Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, vol. I, pp. 412-415, June 2000.
[6] A. Bagnall, J. Lines, A. Bostrom, J. Large, E. Keogh, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," Data Mining and Knowledge Discovery, pp. 606–660, 2017.
[7] A. Bagnall, J.Lines, W.Vickers and E. Keogh, "The UEA & UCR Time Series Classification Repository," www.timeseriesclassification.com.
[8] Y. Chen, et al. "The UCR Time Series Classification Archive, 2015, www.cs.ucr.edu/~eamonn/time_series_data
[9] J. Lines, A. Bagnall, "Time series classification with ensembles of elastic distance measures," Data Mining and Knowledge Discovery, pp. 565–592, 2015.
[10 P. Schafer, "The boss is concerned with time series classification in the presence of noise," Data Mining and Knowledge Discovery, vol. 29, no. 6, pp. 1505–1530, 2015.
[11] A. Bagnall, J. Lines, J. Hills, A. Bostrom, "Time-series classification with COTE: The collective of transformation-based ensembles," International Conference on Data Engineering, pp 1548–1549, 2016.
[12] B. Lucas, A. Shifaz, C. Pelletier, L. O'Neill, N. Zaidi, B. Goethals, F. Petitjean, G. Webb, "Proximity Forest: An effective and scalable distance-based classifier for time series," Data Mining and Knowledge Discovery, pp. 851–881, 2018.
[13] W. Zhiguang, Y. Weizhong, T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," International Joint Conference on Neural Networks (IJCNN), 2017.
[14] A. Ukil, S. Bandyopadhyay, A. Pal, "DyReg-FResNet: Unsupervised Feature Space Amplified Dynamic Regularized Residual Network for Time Series Classification," International Joint Conference on Neural Networks (IJCNN), 2019.
[15] P. Malhotra, V. TV, L. Vig, P. Agarwal, G. Shroff, "TimeNet: Pre-trained deep recurrent neural network for time series classification," European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, pp 607–612, 2018.
[16] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," NIPS, pp:582-588, 1999.
[17] M Hayes, "Statistical Digital Signal Processing and Modeling," Wiley, 1996.
[18] T. C. O'Haver, "Pragmatic Introduction to Signal Processing 2019: Applications in scientific measurement," 2018. Accessed at: https://terpconnect.umd.edu/~toh/spectrum/IntroToSignalProcessing.pdf
[19] N. Cristianini, J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods," Cambridge University Press, 2000.
[20] S. Salvador and P. Chan, "FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space," 2004. Accessed: https://pdfs.semanticscholar.org/05a2/0cde15e172fc82f32774dd0cf4fe5827cad2.pdf
[21] W. Liu, P.P. Pokharel, and J.C. Principe, "The Kernel Least-Mean-Square Algorithm," IEEE Transactions on Signal Processing, pp. 543–554, 2008.
[22] P.S.R. Diniz, "Adaptive IIR Filters," Adaptive Filtering. The Springer International Series in Engineering and Computer Science, pp. pp 377-436, 1997.
[23] A. Ukil, I. Sahu, C. Puri, A. Mukherjee, R. Singh, S. Bandyopadhyay, A. Pal, "AutoModeling: Integrated Approach for Automated Model Generation by Ensemble Selection of Feature Subset and Classifier," International Joint Conference on Neural Networks (IJCNN), 2018.