On Analyzing the Household Energy Consumption Detection for Citizen Behavioral Analysis Carbon Footprint Awareness by Deep Residual Networks

Arijit Ukil[a], Antonio J. Jara[b] and Leandro Marin[c]

[a] *TCS Research and Innovation, Tata Consultancy Services, Kolkata, India*
[b] *University of Applied Sciences Western Switzerland (HES-SO) Switzerland*
[c] *University of Murcia, Murcia, Spain*

Abstract
With the exponential growth of household activities particularly due to the lock-down in COVID-19 pandemic as well as with the usual trend of amplified use of energy consuming appliances, household energy usages are becoming extremely high. Consequently, high energy consumption pattern results in severe increase of air pollution and carbon footprint. Carbon footprint is mainly caused by the greenhouse gases while burning of fossil fuels for producing different forms of energy. In order to restrict the carbon footprint, one of the approaches is to analyze the citizen behavioral pattern by detecting the household appliances. We propose deep neural network based supervised learning algorithm that is capable of classifying the household appliances from energy consumption data. More specifically, we use deep residual networks (ResNet) where learning of the residual functions makes the trained model more robust by transforming the representation learning problem to residual learning problem. Our empirical study on publicly available relevant datasets from UCR timeseries archive demonstrates significantly better and consistent performance over baseline algorithms and state-of-the-art methods.

Keywords 1
Deep Learning, time series, sensor, classification, residual networks, energy data, carbon footprint, appliance detection

1. Introduction

Global warming and adverse climatic change are supposed to be irreversible and affecting human life to a larger extent. Carbon di-oxide (CO_2) is a greenhouse gas and it is one of the primary reasons of global warming. CO_2 emission restriction is the need of the hour and individual citizen has to be taken the required onus for controlled usage of appliances. Household appliances like refrigerator, washing machine, kitchen appliances, computing device consume lots of energy,

which is produced from burning fossil fuels. Hence, carbon footprint reduction is an inevitable action which is to be predominantly taken up by various Governments and other associations [1]. Under the current lock-down in COVID-19 pandemic, household electricity consumption has also increased to a larger extent. In order to reduce the carbon footprint of a nation, different associations along with Government agencies attempt to understand the appliance usage pattern from individual household. Such analysis is performed over energy data like individual house smart meter readings. The household appliance usage can be

linked to enable dynamic energy consumption charges as well as for inculcating the awareness among the citizens of their individual carbon footprints. We find that different human-centric applications are proposed for remote healthcare are proposed in literature [12, 13, 16]. In this paper, our focus is to macro-level human benefit like carbon-footprint reduction.

The technology advancements have led us to different breakthrough applications and developments. The challenge of detecting the household appliance in a non-intrusive way needs to be done by 1. A strong analytics algorithm and 2. A smart infrastructural support that collects the data from household smart meter and enables the provision for analysis and feedback. In this paper the we assume that the smart infrastructure facility is supported by Internet of Things (IoT) backbone. In this work, our main focus is to develop a strong analytic solution which is required for the purpose of analyzing and detecting the household appliances from smart meter timeseries data. We need to keep in mind that the household energy consumption data from accessing the smart energy meter is sensitive in nature as it reveals the in-house activities. It is felt that appropriate security and privacy infrastructure is required to be implemented [14, 15, 17, 18, 19] and should be made part of the complete analytics eco-system.

We propose deep residual network based supervised learning method to classify different household appliances. The current trend of supervised learning by deep neural networks have demonstrated success of deep residual learning, particularly in the applications of 2D (image) and 3D (video) analytics, mostly for computer vision applications. It is perceived that deep residual learning elegantly solves the menacing learning degradation problem especially when the deep network architecture has good number of layers [2]. With the evidences supporting towards deep residual network as a candidate architecture, we propose deep residual architecture for household appliance detection problem using energy signals. It is to be noted that the deep residual network is largely used in visual analytics applications with 2D or 3D data. The current problem is supervised learning over 1D time series. In this paper, we further use regularization of the network parameters (weights) such that the deep neural network does not overfit with the training datasets.

Unlike, computer vision application, which often enjoys the luxury of millions of training examples, the energy consumption data with associated annotations are very tiny in number. In fact, the collection, annotation and distribution of such data is an expensive process. Owing to the scarcity of the training examples, we feel that an appropriate regularization technique is required to optimally fit the network to the training examples.

Traditionally, time series supervised learning baseline algorithm is the dynamic time warping (DTW) based similarity measures with k-nearest neighbor (kNN) based classifier (DTW-1NN) [3, 4], which is a good benchmark. Symbolic Aggregate Approximation (SAX) is a symbolic representation for time series for dimensionality reduction [5] and sliding window-based SAX with cosine similarity based supervised learning technique SAX-VSM has also provided much needed momentum to time series classification solutions. With the advent and success of multi-layer perceptron or MLP algorithms have been studied by researchers for similar time of classification problems [7]. In this work, we consider DTW-1NN, SAX-VSM and MLP as the relevant baseline algorithms.

We present empirical evidence of the proposed deep residual networks, tailored for energy data analysis through experimentations over publicly available UCR time series archive [7 – 8]. It is observed that our method conveniently performs better than the relevant baseline algorithms.

2. Proposed deep residual network architecture

Deep residual network [2] provides the layer-wise recursive learning (with the basic transformation and layer mapping process is shown in Fig. 1) of $\mathcal{H}_{l+1}(X) = \mathcal{H}_l(X) + \mathcal{G}_l\left(\mathcal{H}_l(X)\right)$, where $\mathcal{G}_l$ is the non-linear neural network (in our method, it is a convolution network), $\mathcal{H}_l(X)$ is the desired mapping at lth layer and the initial condition $\mathcal{H}_0(X) = 0, \mathcal{G}_0\left(\mathcal{H}_0(X)\right) = X$, X is the input time series, which is defined as: $X = [x_1, x_2, x_3, \ldots, x_T]$ be the univariate time series, where $X \in \mathbb{R}^{\mathcal{T}}$ and X is a time series energy consumption signal.

Training data $\mathcal{D}$ consists of consists of N number of time series signals each of length $\mathcal{T}$ and each of the training instances has corresponding class label $L^n \in [1, C], \ C \in \mathbb{Z}$ and $\mathbb{L} = [L^1, L^2, L^3, \dots L^N], n = 1, 2, 3 \dots, N$. Thus the complete training dataset is collection of pairs (X^n, L^n), where $\mathbb{X} = [X^1, X^2, X^3, \dots X^N]$, $X^n \in \mathbb{R}^{\mathcal{T}}$, L^n is the corresponding class labels, $\mathcal{D} = [\mathbb{X}, \mathbb{L}]$. The learning algorithm constructs a function $F: \mathbb{R}^{\mathcal{T}} \to \{1, 2, \dots, C\}$. The learning algorithm requires the (training) dataset $\mathcal{D}$ and generates trained model M. The learning algorithm is further a function of regularization factors ψ and functions Υ along with a collection of necessary hyperparemeters Θ for constructing the trained model M and trained model is generated as: $M \xrightarrow{F: f(\mathcal{D}, \psi, \Upsilon, \Theta)} \hat{L}$

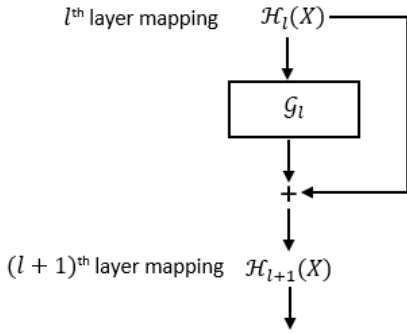$\hat{L} \in [1, C]$ is the predicted inference out.

Figure 1: Basic transformation and layer mapping in deep residual network model.

The individual layers in residual networks modify the learnt representation from the previous layers to counter the vanishing gradient problem [9]. We further find that $\mathcal{H}_{l+1}(X)$ is an additive outcome unlike the conventional deep neural network where transfer function is multiplicative. The underlaying mapping at lth layer $\mathcal{H}_l(X)$ and casts to $\mathcal{H}_l(X) + \mathcal{G}_l(\mathcal{H}_l(X))$ with $\mathcal{G}_l(\mathcal{H}_l(X)) = \mathcal{H}_{l+1}(X) - \mathcal{H}_l(X)$ being the residual function. It is hypothesized that [2] the optimization of the residual mapping is easier than the optimization of the unreferenced raw mapping. Owing to the justification made by the authors and the supported evidences of superior performance of ResNet, we consider that such type of deep residual network is a prudent deep neural architecture choice. Our deep neural architecture is shown in Fig. 2. It consists of three residual blocks, each of the

residual blocks contain number of batch normalization layers along with convolution layers followed by Rectified Linear Unit (ReLU) activation function. Finally, a fully connected dense layer is placed. The final discrimination layer is the softmax function that predicts the output label $\hat{L}$. The predicted label $\hat{L}$. and actual class label L are compared by a loss function (cross-entropy). In this case, we minimize the cost function J over the training examples $\mathbb{X}$ consisting of N number instances, while the is formed by using the stochastic gradient descent algorithm. Given the possibility of insufficiency in the number of training examples, there exists a perpetual possibility of constructing an over-complex model with very high number of network parameters, in terms of weight parameters ω, which is likely to be overfitted on the training distribution without attempting to approximate the source data generation function or the target function. In our earlier work [10], we have proposed strongly regularized convolution neural network SRDCNN for time series classification tasks, which shows the positive impact of regularized learning. Similarly, in this work, we control the deep network parameters by regularization techniques [11]. The proposed deep neural model minimizes the cost function J to find a regularized cost function $\hat{J}$. The regularized cost function is denoted as:

$$\hat{J}(\omega; \mathbb{X}, \mathbb{L}) = J(\omega; \mathbb{X}, \mathbb{L}) + \alpha \Omega(\omega) \quad (1)$$

Where Ω is the regularization function and the regularization factor is $\alpha \in [0, \infty]$.

In this paper, we use the network parameter (ω) norm penalties as expressed above particularly L^2 and L^1 regularizations [11].

We incorporate L^2 regularization (Tikhonov regularization) as:

$$\hat{J}(\omega; \mathbb{X}, \mathbb{L}) = J(\omega; \mathbb{X}, \mathbb{L}) + \alpha \frac{\omega^{\mathrm{T}} \omega}{2} \quad (2)$$

Where, the network parameter gradient is:
$$\nabla_\omega \hat{J}(\omega; \mathbb{X}, \mathbb{L}) = \alpha \omega + \nabla_\omega J(\omega; \mathbb{X}, \mathbb{L}) \quad (3)$$

Subsequently, the weights are updated as:
$$\omega \leftarrow (1 - \epsilon \alpha)\omega + \alpha \nabla_\omega J(\omega; \mathbb{X}, \mathbb{L}) \quad (4)$$

Where, ϵ is the learning rate. From equation (4), it is noted that the weight decay term $(1 - \epsilon \alpha)$ that controls the overall weight vector. Similarly, the Lasso or L^1 regularization is defined as [11]:

$$\hat{J}(\omega; \mathbb{X}, \mathbb{L}) = J(\omega; \mathbb{X}, \mathbb{L}) + \alpha\|\omega\|_1 \quad (5)$$

The following network parameter gradient becomes:

$$\nabla_\omega \hat{J}(\omega; \mathbb{X}, \mathbb{L}) = \alpha \text{sign}(\omega) + \nabla_\omega J(\omega; \mathbb{X}, \mathbb{L}) \quad (6)$$

From equation (3) and equation (6), we are able to note that L^2 and L^1 regularizations impact the network parameters differently. While, Lasso or L^1 regularization attempts to generate sparser weight matrix, Tikhonov or L^2 regularization clips or controls the network weight (ω) values.
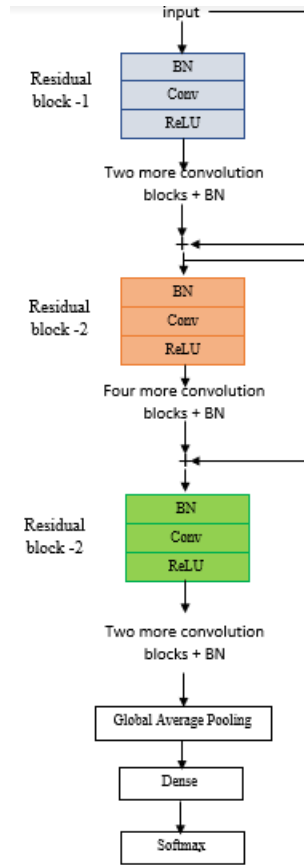

Figure 2. Deep residual network architecture for energy consumption data analytics to detect household appliances. It consists of three consecutive residual blocks along with other required layers.

The hyperparameter set is described in Table 1. One of the noticeable observations is that the network is thinner at the initial and final residual blocks with three convolution layers with number of feature maps of 64 at each layer, while the middle residual block is deeper with five convolution layers with number of feature maps of 128 at each of the layers. From the understanding of the machine learning problem

we attempt to solve, the regularization factor settings play an important role for constructing an effectively learned model. Accordingly, we set the L^1 regularization factor hyperparameter (α_1) to be lower than that of L^2 regularization factor hyperparameter (α_2) with the intent of having lesser sparser weight vectors while the weight vector values are clipped or controlled.

Table 1
Hyperparameter description

Parameter	Brief explanation	Value/ Type
Epoch	Number of times the entire training dataset is iterated	1000
Optimizer	Adaptive learning rate optimization	Adam
Batch size	Number of training samples in one pass	$min\left(\left\|\frac{\mathcal{T}}{10}\right\|, 16\right)$ where $\mathcal{T}$ is the number of sample points at each instant
Number of residual blocks	Total number of residual layers	3
Number of convolution layers at each residual block	Residual block #1	3
	Residual block #2	5
	Residual block #3	3
Kernel size	Residual block #1	{8, 5, 3}
	Residual block #2	{8, 7, 6, 5, 4, 3}
	Residual block #3	{8, 5, 3}
Number of filters	Residual block #1	{64, 64, 64}
	Residual block #2	{128, 128, 128, 128, 128}
	Residual block #3	{64, 64, 64}
α_1	L^1 regularization factor	0.01
α_2	L^2 regularization factor	0.10

3. Experimental Analysis and Results

We consider UCR time series archive with representative datasets which are aligned to the problem statement. The dataset description is made in Table 2. There are five different types

of energy consumption dataset are used for the experimentation purposes. Each of the datasets consists of separate training and testing parts. Our model is first trained over the training dataset and the trained model is tested on the given testing dataset. We report the test accuracies. The experimental datasets represent different types of appliances like kitchen appliances, computing devices and others. UK Government's initiative called 'Powering the Nation', where the behavioral analysis about the usage of electricity by the citizen is used to make an attempt to reduce the carbon footprint. The number of classes also vary among different datasets. With diverse types of appliance detection problem that these datasets (Table 2) represent, we can fairly justify that the experimental evaluation covers large problem areas of detection of appliances from energy consumption data.

Table 2

Energy data (time series) from UCR archive properties

Dataset	Number of classes	Number of training instances	Number of testing instances
Computers	2	250	250
Electric devices	7	8926	7711
Italy power demand	2	67	1029
Large kitchen appliances	3	375	375
Small kitchen appliances	3	375	375

In Table 3, we depict the experimental results of our proposed method. The test accuracy of our method has significantly higher performance merit over the baseline algorithms like MLP [7], DTW-R1-1NN [3] and SAX-VSM [6]. In fact, out of the total five different use cases, our method outperforms rest of the baseline algorithms. In a relative merit, DTW-R1-1NN and SAX-VSM are the closer competitors. With this supporting empirical evidence, we claim that our proposed deep residual network-based model is an apt choice for energy data analysis to detect the household appliances. We further consider SRDCNN as another state-of-the-art algorithm, which has

demonstrated substantially better efficacy than other state-of-the-art [10]. In comparison with SRDCNN, we observe that our method works better in 80% of the total number of datasets. One of the major differences with SRDCNN is the architecture of the deep neural network: SRDCNN is a convolution neural network architecture and ours is deep residual network.

The performance table (Table 3) clearly indicates that the proposed method provides better learning and inferencing capability of energy consumption data to detect the household appliances.

Table 3

Performance in terms of test accuracy metric of our proposed method and related state-of-the-art time series classification algorithms

Sensor name	MLP [7]	DTW-R1-1NN [3]	SAX-VSM [6]	SRDCNN [10]	Our method
Computers	0.496	0.70	0.620	0.781	0.788
Electric devices	0.641	0.602	0.705	0.707	0.723
Italy power demand	0.946	0.950	0.816	0.955	0.964
Large kitchen appliances	0.480	0.795	0.877	0.852	0.907
Small kitchen appliances	0.333	0.653	0.579	0.795	0.709
Total Count	**0**	**0**	**0**	*1*	**4**

4. Conclusion

Carbon footprint reduction is one of the most important problems for creating awareness drive to understand the carbon footprint of individual households to achieve the goal of manifold reduction of overall carbon footprint. In that regard, we propose an analytic solution to detect the appliances at the households using energy consumption data, which is available from the smart energy meter recording. We propose a robust detection algorithm by using deep residual network along with regularization. Our proposed method has shown considerably better test accuracy over

the baseline algorithms for various appliance detection tasks. This proposed method is part of the larger eco-system that attempts to build a convergent human-centric application for the betterment of all of us. We hope that our analytics method provides the required impetus for such human-centered purposes and the global warming concerns can be addressed through citizen-level awareness.

5. Acknowledgements

6. References

[1] R. Aicheleand G. Felbermayr, "Kyoto And the Carbon Footprint of Nations," *Journal of Environmental Economics and Management* 63, no. 3 pp. 336-354, 2012.

[2] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *CVPR*, 2016.

[3] E. Keogh and C. A. Ratanamahatana, "Exact Indexing of Dynamic Time Warping," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 358–386, 2005.

[4] R.J. Kate, "Using dynamic time warping distances as features for improved time series classification," *Data Min Knowl Disc*, pp. 283–312, 2016.

[5] E. Keogh, J. Lin and A. Fu "HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence," *IEEE ICDM*, pp. 226 - 233, 2005.

[6] P. Senin, and S. Malinchik, "SAX-VSM: Interpretable Time Series Classification Using SAX and Vector Space Model," *IEEE ICDM,* pp.1175,1180, 2013.

[7] A. Bagnall, J. Lines, W. Vickers and E. Keogh, "The UEA & UCR Time Series Classification Repository," *www.timeseriesclassification.com*

[8] A. Bagnall, J. Lines, A. Bostrom, J. Large and E. Keogh, "The Great Time Series Classification Bake Off: A Review and Experimental Evaluation of Recent Algorithmic Advances," *DMKD*, 2017.

[9] Y. Shen, J. Gao. "Refine or Represent: Residual networks with explicit channel-wise configuration," *IJCAI*, 2018.

[10] A. Ukil, A. Jara, L. Marin, "SRDCNN: Strongly Regularized Deep Convolution Neural Network Architecture for Time-series Sensor Signal Classification Tasks," *arXiv:2007.06909*, 2020.

[11] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," *The MIT Press*, 2016.

[12] C. Puri, A. Ukil, S. Bandyopadhyay, R. Singh, A. Pal, and K. Mandana, "iCarMa: Inexpensive Cardiac Arrhythmia Management--An IoT Healthcare Analytics Solution," *IoT-Enabled Healthcare and Wellness Technologies and Systems Workshop*, 2016.

[13] A. Ukil, S. Bandyopadhyay, C. Puri, and Arpan Pal, "Heart-trend: An Affordable Heart Condition Monitoring System Exploiting Morphological Pattern," *IEEE ICASSP*, pp. 6260-6264, 2016.

[14] A. Ukil, "Privacy Preserving Data Aggregation in Wireless Sensor Networks," *IEEE International Conference on Wireless and Mobile Communications*, pp. 435-440, 2010.

[15] A. Ukil, "Security and Privacy in Wireless Sensor Networks," *INTECH Open Access Publisher*, pp. 395- 418, 2010.

[16] A. Ukil, A. J. Jara, and L. Marin, "Data-Driven Automated Cardiac Health Management with Robust Edge Analytics and De-Risking," *Sensors*, June, 2019.

[17] A. Ukil, "Secure Trust Management In Distributed Computing Systems," *IEEE International Symposium on Electronic Design, Test and Application*, 2011.

[18] T. Bose, S. Bandyopadhyay, A. Ukil, A. Bhattacharyya and A. Pal, "Why Not Keep Your Personal Data Secure Yet Private In IoT?: Our Lightweight Approach," *IEEE ISSNIP,* 2015.

[19] A. Bhattacharyya, T. Bose, S. Bandyopadhyay, A. Ukil, and A. Pal, "LESS: Lightweight Establishment of Secure Session: A Cross-Layer Approach Using CoAP and DTLS-PSK Channel Encryption," *IEEE International Conference on Advanced Information Networking and Applications Workshops,* pp. 682-687, 2015.