2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 11-14, 2020. Toronto, Canada

Light Weight Dilated CNN for Time Series Classification and Prediction

Pranav Khanna
Department of Electrical Engineering
Indian Institute of Technology, Indore
Indore, India
pranikhanna1998@gmail.com

Apurva Narayan
Department of Computer Science
The University of British Columbia
BC, Canada
apurva.narayan@ubc.ca

***Abstract*—Time series data is available from a diverse set of sensors in real life. It is of prime importance in the domain of machine learning and artificial intelligence to analyze such data and identify outliers or anomalies, characteristic of the underlying activities and predict the future. Traditionally, time-series analysis involves identifying features using exploratory data analysis and using statistical approaches for classification and prediction. However, with the advent of convolutional neural networks (CNN), our ability to extract features automatically has substantially improved. In this paper, we propose a novel light-weight deep learning architecture of dilated CNN for classification and predicting time series data sets. We evaluate our model on a real-world *human activity recognition* time series data set and a synthetically crafted pseudo-realistic dataset for *human intent recognition*. Our model outperforms the state-of-the-art models and is light-weight.**

I. Introduction

Today sensors are ubiquitous and can be found in almost any technical device ranging from washing machines to pacemakers and wearable devices. These sensors produce large quantities of data in the form of time series. Time series analysis finds application in a wide range of fields such as health care [1], climate [2], robotics [3], and finance [4]. It is of high value to consumers and manufacturers to leverage the large quantity of data in real-time to make useful inferences from it. In the domain of machine learning and artificial intelligence, we want to use these time-series signals and either perform classification or regression tasks that help us identify system behaviour, makes them safe, secure and resilient. There is substantive evidence of machine learning research in the domain of time series classification [5]. However, either these techniques cannot handle big data or are not accurate for the task, and the models are computationally expensive that cannot execute on low power devices.

In this paper, we propose a novel light weight deep learning architecture of dilated CNN (dCNN) for time series classification. Our work adapts the dCNN architecture from [6] where authors introduce the concept of dilations and its effects on the convolution operation. It highlights the benefits of dCNN for performing semantic segmentation. We evaluate the performance of our model on popular activity recognition datasets Mobifall [7] and WISDM [8]. Moreover, we also highlight the computational efficiency of dCNN when compared to the state of the art models consisting of LSTMs and GRUs. We believe that this lightweight architecture could pave the way for running highly accurate models on embedded systems for various applications and would reduce the need for expensive systems.

The second major contribution of this paper is the introduction of the novel dataset for Human Intent Recognition (HIR) or Activity Prediction. HIR is an upcoming field of research which aims to recognize the intent of a user based on a sequence of actions. HIR extends Human Activity Recogntion (HAR) by bringing time factor into consideration by predicting the final action of a user that might happen in the future, unlike HAR, which recognizes an activity after seeing the data. We introduce a novel pseudo-realistic dataset created from an existing HAR dataset [8] for HIR. The dataset has been carefully developed such that it reflects practical situations. Each sample in the dataset combines activities from the HAR dataset in different combinations which would help predict the final intent of the user. We benchmark the performance of our dCNN model, LSTM and GRU on this new dataset. These models are evaluated based on computational complexity and predictive accuracy.

II. Related Work

Analyzing data from sensors is same as performing time series analysis. It is a well-studied problem and still remains the central component in diverse areas of research. For instance, in [9] they combined the statistical machine learning approach of auto-regressive and moving average with artificial intelligence (AI) techniques to provide accurate time-series predictions. In AI, recurrent neural networks are traditionally considered a de-facto for modelling time series data. In [5] used a recurrent neural network for time series forecasting whereas as [10] use the modern variant of recurrent networks, the long-short term memory networks for forecasting. Time series forecasting has been an active area of research for deep learning researchers as well as where [11] and [12] employed deep recurrent and deep convolutional neural networks for time series classification. There is also extensive research in computational intelligence techniques such as fuzzy logic [13] and Bayesian models [14] for time series forecasting.

Deep neural networks have found profound success in the domain of computer vision and natural language processing with advances in convolutional and recurrent neural networks. However, recently, researchers are investigating the use of various variants and state-of-the-art deep learning models in the domain of time series analysis. In [15] they use the WaveNet model for multivariate time series prediction where conditional dependence of each variate on the other is given special consideration. In another work [16] designed a multi-channel CNN where each channel handles a variable in a multivariate time series. Time series classification plays a key role in numerous industrial applications such as in [17] use a CNN for anomaly detection and diagnosis in semiconductor manufacturing. In [18] presented a novel architecture of time series classification with each time series transformed to a 3-D tensor and subsequently processed through the network. A comprehensive survey on deep learning techniques for time series forecasting is presented in [19], and a robust benchmarking of algorithms is presented in [20]. Time series from the domain of HAR has been studied in [21] where CNN is applied on the frequency features of a time series.

Past research has explored the domain of HIR and is sometimes referred to as plan recognition and goal recognition. In [22] uses goal recognition to recognize a player's high-level intentions using a computational model trained on a player behaviour corpus. The paper highlights the advantages of deep LSTM based goal recognition models over single-layer LSTMs, n-gram encoded feedforward neural networks and Markov logic networks. In [23] authors have applied plan recognition on the popular game Starcraft navigation and Monroe Plan Corpus. The paper presents a recursive neural network model that learns such a decision model automatically. In [24] recursive Bayesian approach has been used for HIR in the context of shared control robotics. It is fused with multiple non-verbal observations to probabilistically explain the intent of the user. The paper concludes that its approach outperforms existing solutions and demonstrates that it improves intent inference and performance for shared control operation. In [25] developed HIR model in the domain of Natural Language Processing (NLP) to retrieve relevant information from a sentence for a voice-based human-machine interface for modern intelligent vehicles. In another work [26] propose a new model to predict intention of neighbouring vehicles from raw sensor data for self-driving cars.

III. Data Set and Experiment

To evaluate the performance of our proposed model, we apply the model to two publicly available datasets on human activity recognition containing accelerometer data. We subsequently propose a pseudo-synthetic dataset for human intent recognition adapted from the Wisdm HAR dataset. The results are compared against the published results on these datasets.

A. Human Activity Recognition datasets

1) Mobifall Dataset
: The Mobifall dataset [7] contains data for three sensors. An accelerometer, a gyroscope and a software-based orientation sensor. A Samsung Galaxy S3 device is used to capture the motion data. The accelerometer sensor gives the acceleration force, the gyroscope sensor gives the rate of rotation and the orientation sensor gives the angle around the x, y and z axes.

The dataset consists of activities related to daily living like standing, walking, jogging, jumping, moving up the stairs, moving down the stairs, sitting on the chair, stepping in the car and stepping out of the car.

2) Wisdm Dataset
: The Wisdm dataset [27] contains data for the sensor, accelerometer. The dataset was collected from users using an Android smartphone who carried the phone in their trousers and asked to walk, jog, ascend stairs, descend stairs, sit, and stand for specific periods of time. The accelerometer data was collected at a rate of 20 samples per second. The accelerometer sensor in this case also gives acceleration force along the x, y and z axes (including gravity). The acceleration recorded includes gravitational acceleration toward the centre of the Earth. The dataset includes the activities walking, jogging, upstairs, downstairs, sitting, standing.

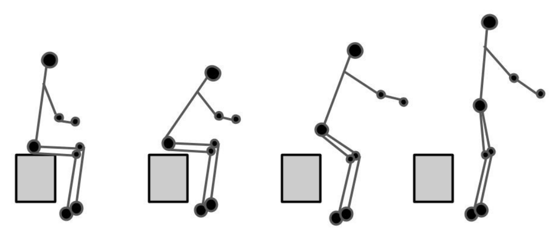

Fig. 1. Positions under consideration for the Intent Sitting to Standing. The second and third positions correspond to the Transition positions for the Intent Sitting to Standing

B. Human Intent Recognition Dataset

The HIR dataset includes the intents from sitting to standing, standing to walking, and walking to jogging. The Wisdm HIR dataset is split into a training dataset containing data from 11 users and a test dataset containing data from 4 users.

$$A = Sensor\ values\ for\ Primary\ Position : A \tag{1}$$
$$B = Sensor\ values\ for\ Primary\ Position : B \tag{2}$$
$$T = Sensor\ values\ for\ Transition\ Position \tag{3}$$
$$S = Sample\ of\ the\ HIR\ dataset \tag{4}$$
$$T = A \cap B \tag{5}$$
$$K1 = Total\ Number\ of\ timestamps\ for\ which\ data\ for\ position\ A\ is\ recorded\ in\ a\ sample. \tag{6}$$
$$K2 = Total\ Number\ of\ timestamps\ for\ which\ data\ for\ Transition\ position\ is\ recorded\ in\ a\ sample. \tag{7}$$
$$K3 = Total\ Number\ of\ timestamps\ for\ which\ data\ for\ position\ B\ is\ recorded\ in\ a\ sample. \tag{8}$$
$$K1 = [1, 8],\ K2 = [1, 5],\ K3 = [0, 8] \tag{9}$$
$$S = \{A_{n=0},\ A_{n=1},\ A_{n=2},\ .....\ A_{n=k1-1},\ A_{n=k1},\ T_{n=k1+1},\ T_{n=k1+2},\ T_{n=k1+3},\ .....\ T_{n=k1+k2-1},\ T_{n=k1+k2},\ B_{n=k1+k2+1},\ B_{n=k1+k2+2},\ B_{n=k1+k2+3},\ .....\ B_{n=k1+k2++k3-1},\ B_{n=k1+k2+k3}\} \tag{10}$$

1) Transition Positions

: For every intent, there are three positions taken into consideration in this paper, two primary positions and one transition position. These transition positions help us identify the final intent of the person. The movement from one primary position to another primary position is a smooth continuous movement. The intermediate positions between the two primary positions are collectively grouped as transition positions for that intent. The overlapping sensor values of the two primary positions are considered to be the values for transition positions. Consider the positions in the Figure 1 which illustrate the different positions taken into consideration for the intent *Sitting* to *Standing*. The intermediate positions between the two primary positions are collectively defined as transition positions for that intent.

2) Procedure

: Equations 1 - 10 give an insight about the process of creating the HIR dataset.

1) Equation 1, 2 and 3 represent the sensor values for different positions. Equation 5 illustrates the method of obtaining the sensor values for the transition position. The overlapping values of the two primary positions in each dimension x, y, and z are considered as the sensor values for the transition position.
2) Equation 6, 7, 8 describe the features of the HIR dataset sample. Consider a sample with {*K1 = 3, K2 = 5 , K3 = 4*}. The first three timestamps of the sample contains the data for position A , the next five timestamps for position T and the last four timestamps for position B. The main feature of the HIR is its ability to predict the intent irrespective of the number of events happening before and after the transition and different values of K1, K2 and K3 help reflect practical situations in the HIR dataset samples.
3) A special case arises when *K3 = 0*, which highlights the main difference between HAR and HIR. Thus when K3 = 0, the sample does not contain the sensor data for the destination primary position. This highlights that HIR is able to predict the final position of the user on the basis of its starting and transition position, which is however not the case in HAR which is only able to recognise the position after it receives the sensor data for that position. One can refer HIR as HAR with a time lag.

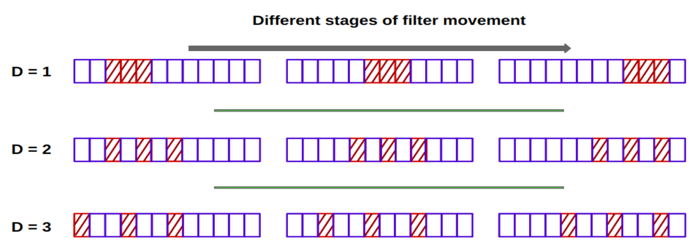

Fig. 2. Different stages of filter movement across the input for different dilation rates in 1d convolutions. The filter in this image is of size three.

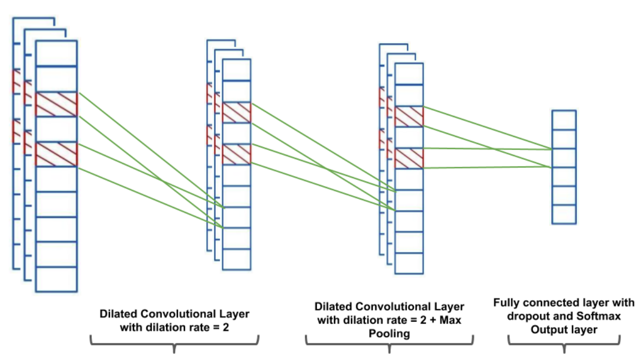

Fig. 3. Block Diagram of the proposed Neural Network Architecture. The red shaded boxes illustrate filter operation on the layer input.

IV. Dilations in CNN

Till recently, Recurrent neural networks (RNNs) have been considered the most optimum neural network architectures for time series analysis however recent studies show that CNNs can provide the same performance as RNNs and may even exceed them. Dilated convolutions help one-dimensional CNNs to effectively learn time series dependencies. CNNs have many advantages over RNN based solutions. They are much more computationally efficient than RNNs and are much more easier to train. Further hidden layers in CNNs can be visualised which is not the case in RNNs and this helps in the explaining the predictions of the model. Figure 2 illustrates the effect of change in dilation rate on filter movement across the input. The filter used for performing convolutions is of size three.

V. dCNN Module

Our proposed approach uses the dCNN Module. The dilated convolutional neural network (dCNN) module is composed of a sequence of layers of dilated convolutional layers with ReLU (Rectified Linear Unit) activations. We apply a 1-dimensional dilated convolution to extract features from the segments of the time series. In the dCNN architecture presented in Figure 3, the first layer in the dCNN module is the dilated convolutional layer; filters in this layer are of $size = 2$ and slide over the sensor time series with $stride = 1$ and $dilation\ rate$ $= 2$ in the vertical direction. Moreover, employing dilations in the convolution layer allows one to learn features between the observations that are far off.

The next layer in the dCNN module is similar to the first layer that performs dilated convolutions on the feature maps obtained from the previous layers. It allows one to extract features between much far off sensor values. Subsequently, we have a max pooling layers that is used to abstract the high-level features that are learnt by our dCNN filters. Finally, we have the last fully-connected layer that is applied to the outputs of the dCNN layers. The softmax layer computes the probability values for the predicted class.

A. dCNN Architecture Analysis

The network designed for the task of human activity recognition and prediction has four layers including: Dilated Convolution Layer 1 (No. of filters = 8, Filter Size = 2, Dilation Rate = 2) $\Longrightarrow$ Dilated Convolution Layer 2 (No. of filters = 8, Filter Size = 2, Dilation Rate = 4) $\Longrightarrow$ Max Pooling Layer and followed by a fully connected layer. To train the network, we used batch size of 2000 on Google Cloud Instance with 1 NVIDIA Tesla T4 GPU. We used the $cross - entropy$ loss function with the ADAM optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and the learning rate was set to 10^{-5}. We used drop out at a rate of 0.5.

B. LSTM and GRU Architecture Analysis

We develop two architectures of single layer network of LSTMs and GRUs to compare the effectiveness of our dCNN model. We use a batch size of 2000 to train each network and use ADAM optimizer with hyperparameter $\beta_1 = 0.9$ and $\beta_2 = 0.999$. A softmax layer is the terminating layer in each of the architecture for classification.

VI. Results

In this Section we present our results on the three datasets. Two for activity recognition and one for pseudo-synthetic dataset for activity prediction or intent recogntion. We compare state of the art methods with our approach and demonstrate superiority in performance along with the model being light weight enabling deployment on real-world low power mobile devices.

We present our results on activity recognition. In Table I we present our results for the Wisdm dataset on the recurrent neural networks based on LSTM, GRU, and our dCNN model. We compare the performance of our model on accuracy and size of the model (the number of parameters in the model).

Model Name	Accuracy	# of Parameters
LSTM	96.73%	94,214
GRU	96.12%	89,862
dCNN	**94.95%**	**9,606**

TABLE I

Performance evaluation of the dCNN model on the WISDM data set for activity recogniton

Similar to the results on the Wisdm dataset, we performed experiments on another publicly available dataset, the Mobifall, to evaluate the generalizability of our model. It turns out that results are promising and reveal interesting insights about the process. In Table II we present our results for the Mobifall dataset and compare the results obtained from state-of-the-art models.

Model Name	Accuracy	# of Parameters
LSTM	81.05 %	132,617
GRU	80.98%	128,265
dCNN	**78.16%**	**7,257**

TABLE II

Performance evaluation of the dCNN model on the Mobifall data set for activity recognition

In both Table I and Table II the results follow a similar trend where the dCNN model has better or nearly as good as state-of-the-art performance however, the model has significantly low number of parameters. The low number of parameters make the model modular and portable in terms of devices with low memory and power.

We briefly introduced the notion of intent recognition as a problem similar to predicting the activity instead of classifying it ahead of time. We created a pseudo-synthetic data set for researchers interested in the problem of intent recognition.In Table III, we present results and comparison of the dCNN model for HIR. Our results are state-of-the-art at the moment. Our dataset is opensource and we welcome researchers to test their models and present results in the domain of human intent recognition.

Model Name	Accuracy	# of Parameters
LSTM	98.65 %	22,788
GRU	98.74%	18,436
dCNN	**96.259%**	**452**

TABLE III
PERFORMANCE EVALUATION OF THE DCNN MODEL ON THE INTENT RECOGNITION DATASET

It may be noted from the results in Table III that although the model's performance is approximately 2% lower however there is a 40-fold reduction in the model size. It is a very useful result and paves way forward for real-time intent recognition algorithms that can be use in health care and automotive industry.

VII. SUMMARY

We have proposed a new convolutional neural network architecture for sensor data classification and prediction. We present a novel dilated CNN model for human activity recognition and human intent recognition. The dCNN architecture uses stacks of dilated convolutions. The extracted features for classification task are obtained by considering the inter and intra relation between variates. Our experiments show that the proposed model is as effective as previous models working with hand-crafted features such as spectrogram and statistical features. In this paper we also introduce a novel dataset of Human Intent Recognition using Human Activity Recognition. Our work paves way for two specific domains, human activity/intent recognition from sensor data and a generic approach using dCNN for time series analysis that has widespread application.

REFERENCES

[1] S. S. Jones, R. S. Evans, T. L. Allen, A. Thomas, P. J. Haug, S. J. Welch, and G. L. Snow, "A multivariate time series approach to modeling and forecasting demand in the emergency department," *Journal of biomedical informatics*, vol. 42, no. 1, pp. 123–139, 2009.

[2] Z. Zhang, Zhang, and Khelifi, *Multivariate Time Series Analysis in Climate and Environmental Research*. Springer, 2018.

[3] C. Pérez-D'Arpino and J. A. Shah, "Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 6175–6182.

[4] N. Maknickienė, A. V. Rutkauskas, and A. Maknickas, "Investigation of financial market prediction by recurrent neural network," *Innovative Technologies for Science, Business and Education*, vol. 2, no. 11, pp. 3–8, 2011.

[5] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, p. 6085, 2018.

[6] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *CoRR*, vol. abs/1511.07122, 2015.

[8] G. M. W. Jennifer R. Kwapisz and S. A. Moore, "Activity recognition using cell phone accelerometers," in *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10), Washington DC*, 2010.

[7] G. Vavoulas, M. Pediaditis, E. G. Spanakis, and M. Tsiknakis, "The mobifall dataset: An initial evaluation of fall detection algorithms using smartphones," in *13th IEEE International Conference on BioInformatics and BioEngineering*, 2013.

[9] G. P. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[10] K. Yeo, "Model-free prediction of noisy chaotic time series by deep learning," *arXiv preprint arXiv:1710.01693*, 2017.

[11] B. P. Orozco, G. Abbati, and S. Roberts, "Mordred: Memory-based ordinal regression deep neural networks for time series forecasting," *arXiv preprint arXiv:1803.09704*, 2018.

[12] O. Yazdanbakhsh and S. Dick, "Multivariate time series classification using dilated convolutional neural network," *arXiv preprint arXiv:1905.01697*, 2019.

[13] ——, "Forecasting of multivariate time series via complex fuzzy logic," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 2160–2171, 2017.

[14] F. A. Saad and V. K. Mansinghka, "Temporally-reweighted Chinese restaurant process mixtures for clustering, imputing, and forecasting multivariate time series," in *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS 2018)*, ser. Proceedings of Machine Learning Research, vol. 84. Playa Blanca, Lanzarote, Canary Islands: PMLR, 2018, pp. 755–764.

[15] A. Borovykh, S. Bohte, and C. W. Oosterlee, "Conditional time series forecasting with convolutional neural networks," *arXiv preprint arXiv:1703.04691*, 2017.

[16] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, "Exploiting multi-channels deep convolutional neural networks for multivariate time series classification," *Frontiers of Computer Science*, vol. 10, no. 1, pp. 96–112, 2016.

[17] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, no. 2, pp. 135–142, 2017.

[18] C.-L. Liu, W.-H. Hsaio, and Y.-C. Tu, "Time series classification with multivariate convolutional neural network," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4788–4797, 2018.

[19] J. C. B. Gamboa, "Deep learning for time-series analysis," *arXiv preprint arXiv:1701.01887*, 2017.

[20] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, 2019.

[21] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "A deep learning approach to on-node sensor data analytics for mobile or wearable devices," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 56–64, 2016.

[22] W. Min, B. W. Mott, J. P. Rowe, and J. C. Lester, "Deep lstm-based goal recognition models for open-world digital games," in *AAAI Workshops*, 2017.

[23] F. Bisson, H. Larochelle, and F. Kabanza, "Using a recursive neural network to learn an agent's decision model for plan recognition," in *IJCAI*, 2015.

[24] S. Jain and B. Argall, "Recursive bayesian human intent recognition in shared-control robotics," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2018, pp. 3905–3912.

[25] Y. Zheng, Y. Liu, and J. H. L. Hansen, "Intent detection and semantic parsing for navigation dialogue language processing," *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, 2017.

[26] S. Casas, W. Luo, and R. Urtasun, "Intentnet: Learning to predict intention from raw sensor data," in *Proceedings of The 2nd Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, A. Billard, A. Dragan, J. Peters, and J. Morimoto, Eds., vol. 87. PMLR, 29–31 Oct 2018, pp. 947–956. [Online]. Available: http://proceedings.mlr.press/v87/casas18a.html

[27] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.