Automatic Time-Lapse Neural Network Modelling Based On Low-Level Data

Rahul Vishnoi
Department of Electronics and Communication Engineering
Faculty of Engineering, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India

ABSTRACT: Automatic process modelling (APM) is an enabling technology for the development of smart fabrication systems (IMSs). The analysis of obtained models enables the prompt identification of error-prone measures and the design of effective mitigation strategies, from parameter optimization to the production of tailored staff training, in all aspects of the manufacturing process. In this work as propose a Time Delay Neural Network (TDNN) applied to low-level data for the automatic recognition of various process phases in collaborative industrial tasks. As selected TDNN because, while retaining computational performance, they are suitable for modelling time dependent processes over long sequences. As acquired two novel datasets reproducing a standard IMS environment to experimentally evaluate the recognition efficiency and the generalization capability of the proposed process. Datasets (including manually annotated ground-truth labels) are publicly accessible to allow other methods to be evaluated on them, replicating a standard environment for Industry 4.0. The first dataset replicates a collaborative robotic system in which a human operator communicates with a robotic manipulator while performing a pick and place function. The second package is a human tele-operated, aided robotic manipulation for assembly applications. The results obtained are superior to other literature methods, and indicate an improved computational efficiency.

KEYWORDS: Intelligent manufacturing, Industry 4.0, Time delay neural network, TDNN, Collaborative Robot.

INTRODUCTION

The emergence of Industry 4.0 (I4) has introduced new modes of production to allow greater flexibility in processes without penalizing cost and performance. According to one of I4's nine implementation pillars is collaborative robots that help human operators manage different stages of the production process. The collaborative model represents one of the development goals to be achieved in I4, achieving the best outcome of quality and performance while optimizing flexibility. Achieving these targets is very difficult because, on the one hand, as need to optimize robotic and development parameters while, on the other, achieving the best human-operator efficiency possible without penalizing ergonomics and user experience [1], [2]. Flexibility is one of the key survival needs in today's competitive markets, particularly for small and medium-sized enterprises (SMEs), but the implementation of collaborative robots encounters challenges related to management and incorporation in the industrial environment. From this the lack of empirical literature on the simulation of collective industrial activities follows. Automatic process modelling (APM) can promote the introduction of collaborative robot in small and medium-sized enterprises by offering a more abstract and user-friendly understanding of collaborative systems and easier introduction with decision taking and intelligent manufacturing systems.

The use of low-level data is consistent with I4 guidelines reflecting a growing proliferation of cyber-physics systems, and providing easy access to heterogeneous data. Automatic collaborative system process analysis enables the timely identification of error-prone measures and the design of effective mitigation strategies [3]covering all aspects of the operation, from parameter optimization to predictive maintenance. Automatic process modelling (APM) can promote the introduction of collaborative robot in small and medium-sized enterprises by offering a more abstract and user-friendly understanding of collaborative systems and easier introduction with decision taking and intelligent manufacturing systems.

The use of low-level data is consistent with I4 guidelines reflecting a growing proliferation of cyber-physics systems, and providing easy access to heterogeneous data. Automatic collaborative system process analysis enables the timely identification of error-prone measures and the design of effective mitigation strategies, covering all aspects of the operation, from parameter optimization to predictive maintenance. TDNNs were successfully applied to the study of heterogeneous data (e.g., kinematic data of low dimensionality, images, and videos). As acquired two novel datasets reproducing an I4 IMS setting in order to experimentally test the recognition efficiency and the generalization capability of the proposed APM process. Publicly accessible data sets (including video recordings, low-level sensor data and ground-truth annotations) to allow benchmarking of methods [4], [5]. The first dataset replicates a collaborative robotic system in which a human operator participates in a pick-and-place role with a robotic handler. The TDNN results are superior to other methods available in literature maintaining a reasonable computational cost. The proposed method and novel datasets are key-components in the development of future IMS with advanced situation awareness capabilities.

METHOD

As will first define our TDNN-based approach then present another three supervised APM algorithms that are used as benchmarks. Various methods and datasets allow us to compare the proposed network's ability with the affine standard approaches.

Time-delay neural network:

The proposed TDNN has a pyramidal structure which gives them a wider temporal context, i.e. the initial transformations are learned on the values of narrow ranges and the deeper layers process the hidden activations from a wider temporal context due to node dilation as shown in Figure 1. Since the convolutions occur in the time axis, stacking layers with and increasing dilatation rate allows the model to learn wider temporal relationships, resulting in a higher abstraction of the feature. Lower layers of the network are modified during back-propagation by a gradient accumulated over all the time steps in the temporal sense of the input.

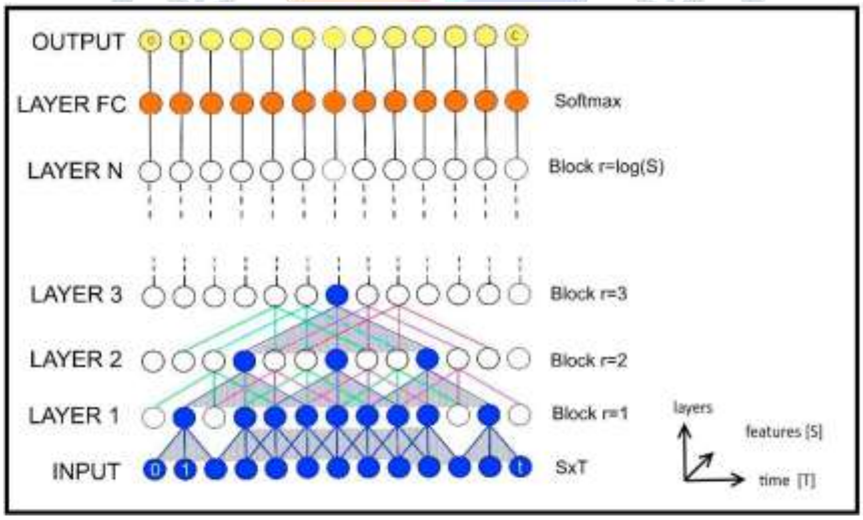

Figure 1. TDNN network architecture

Other methods:

In temporal modelling, as benchmark the proposed TDNN with other 3 methods focused on different approaches: Long-Short Term Memory (LSTM), Support Vector Machine (SVM) and Random Forest (RF) [7]–[8]. Such methods were chosen because, as stated, they prove effective in industrial supervised task modelling.

LSTM architecture is an important part of recurrent neural networks. LSTM has been commonly used in speech recognition and temporal dependence capture using parameters that measure past cell incidence on present state. Back Propagation through Time (BPTT) recalculate the gradient for each step with respect to the weights and sum it up over time stages. In temporal modelling, as benchmark e proposed TDNN with other 3 methods focused on different approaches: Long-

Short Term Memory (LSTM), Support Vector Machine (SVM) and Random Forest (RF). Such methods were chosen because, as stated in, they prove effective in industrial supervised task modelling.

LSTM architecture is an important part of recurrent neural networks. LSTM has been commonly used in speech recognition and temporal dependence capture using parameters that measure past cell incidence on present state [9]. Back Propagation through Time (BPTT) recalculate the gradient for each step with respect to the weights and sum it up over time stages.

Fig. 2. (a) The hardware training console used by one of the students during data acquisition. (b) Example of the Virtual task considered in the VIT-MR dataset: four coloured rings need to be placed in the corresponding peg. (c) ICRT dataset setup.

Virtual Industrial Task Master-slave Robot dataset:

VIT-MR (Virtual Industrial Task Master-slave Robot) dataset was developed to replicate all the steps typical in high precision small-scale manufacturing of robotic assisted tele-operated manipulation processes. The user controls Leo master console remote slave manipulators (BBZ srl, Verona, Italy) seen in Figure 2a, a lightweight hardware system that combines two manipulators of maters, a high-definition stereo monitor and a foot pedal tray. The console ensures an interactive user interface which allows control of ergonomic slave manipulators and enhanced magnified vision. Simulated slave manipulators, visible in Figure 2b provide high dexterity and scaling of movement to ensure accurate and stable manipulation of components during assembly process.

As used a research version of Xron (BBZ srl, Verona, Italy), a realistic virtual simulator suitable for high fidelity applications, such as medical training, to incorporate simulated environments. As used this experimental setup because it is able to replicate kinematic variables very close to a real industrial robotic master-slave environment. The function of manipulation involves putting a set of coloured rings in their proper position on a peg board, as shown in Figure 2b. The exercise consists of raising a ring with one of the robotic arms, moving the ring onto the other robotic arm, and putting the ring in the appropriate position.

In this study was enrolled a group of 17 users with no specific experience in robotic aided manipulation. Both consumers have no prior experience of more than one hour using similar robotic systems. That topic had a one-hour time slot, the first half is devoted to practice with the actual platform, and the second half is devoted to the reported trials. Growing topic conducted from a minimum of ten to a maximum of twenty trials which resulted in a total of 256 sequences. Multiple users with varying levels of experience allow the model to be resilient to the difference in execution of movements and to better identify transitions of phases.

Table 1. Datasets features description.

Index	Description
VIT-MR	
0	Process phase
1-15	Rotation and translation Right
16	Gripping Angle Right
17-28	Rotation and translation Left
29-31	Cartesian position Left
32	Gripping angle Left
ICRT	
0	Gesture
1-7	ArUco Marker Cartesian position and orientation in quaternions. The reference point of the system is given by the RealSense camera [17]
8-16	Robot joint angles
17	Timestamp
18-25	End-effector Cartesian and orientation
26-114	Leap motion hand feature (87) [18]

Table 2. Datasets gesture description.

Index	VIT-MR	ICRT
0	Hand pick the tool with the ring	Colleting the ring
1	Move the ring with the tool at an arbitrary point	Passing the ring from the right arm to the left arm
2	Release the tool	Posing the ring in the correct pole
3	Robot identify and pick the ring	Failing grabbing the ring
4	Let the robot move the ring at drop point	Failing passing the ring from right arm to left arm
5	Robot release the ring	Failing posing the ring in the correct pole

The dataset consists of synchronized stereo images, kinematic variables for the two slave manipulators, and other variables of device status. 16 variables are available for each slave manipulator: Cartesian position, rotation matrix, and angle of grabbing method as shown in Table 1. Those variables represent all of the system's raw data. As ruled out images to avoid pre-processing data. The phases annotated manually are six, and are listed in Table 2. They provide classification of errors that occur during execution of tasks to better explain results unique to the user. The dataset and associated extensive documentation can be found at gitlab.com/altairLab/VIT-MR.git.

Industrial Collaborative Robot-human Task dataset:

ICRT (Industrial Collaborative Robot-Human Task) data set includes the use of four devices for human operator interaction and a collaborative robot interaction. A Leap Motion system (LeapMotion, US), an ArUco marker, an Intel RealSense D415(Intel, US) camera and a Panda robot (Franka Emika, Germany) are the sensors used. As in a typical I4 sense, as decided to simulate an environment that contains multiple sensors. Thus, as used data derived directly from machines (robot cinematics) and data collected from the supervising sensors such as leap motion and ArUco marker. As picked a ring for avoiding problems due to object manipulation such as the robot's incorrect grip. Using a device allows one to expand the experiment to instances where hazardous objects are mounted or are not appropriate for human interaction. The setup is shown in Figure 2c which highlights the sensors used and the task's salient points. The experiment is outlined as follows: in a first step, the ring is selected by a human operator using a device to avoid direct hand contact. The consumer then moves it to an arbitrary location and releases the device. The second step includes the RealSense camera sensor's scene segmentation and ring recognition and then the robot picks the ring and positions it sub sequentially at the drop point, then releases the ring. When the job finishes robot returns to a ready location waiting for a new job to be performed. The Leap Motion data will describe the entire mission, recording hand motions, the location of the robot joints and the end-effector pose to capture the robot's motion, and the ArUco will monitor the ring's motions. These values are labelled in Table 1. The function was performed by a single human operator for 40 times. The data set and related detailed documentation can be found at gitlab.com/altairLab/ICRT.git

EVALUATION AND RESULT

The VIT-MR dataset was evaluated using LOUO (Leave One User Out) methodology which consists of removing and using a user during the training process. That means the preparation and the examination are performed 18 times. As used leave k-out (with kk = 5) methodology for the ICRT dataset, which consists of leaving 5 samples in the training process and using them as samples. As the total number of trials was 40, this results in 8 split training. The values shown reflect the mean values with standard deviation for both datasets. For both datasets, as defined as macro/micro accuracy as metrics to benchmark the different methods. In our work classes the classification mark expected by the model is called process phases. Micro average is measured as the sum of total correct predictions across all classes and is advantageous for classification on unbalanced classes as it aggregates the contributions of all phases to calculate the sum metric in terms of precision while taking into account false positives. On the other hand, for each class macro accuracy is measured as the mean of the true positive values. Table 3 displays the findings obtained by application of the methods to the two datasets. The recorded data are collected trying to optimize the accuracy of microphones.

Table 3. Result for Macro and Micro average accuracy for both datasets reported as mean betasen LOUO results.

Method			**Dataset**			
	VIT-MR			ICRT		
	Micro Average	Macro Average	Time (s)	Micro Average	Macro Average	Time (s)
TDNN	**69.76**	**53.85**	**1042**	**86.95**	**79.04**	**256.7**
Std	± 6.08	± 3.508		± 3.693	± 4.902	
LSTM	60.86	42.35	34455	65.83	51.98	6367
Std	± 8.60	± 3.50		± 13.18	± 15.07	
RFC	62.16	42.7	753.69	86.92	80.32	81.98
Std	± 8.914	± 3.792		± 3.558	± 5.126	
SVM	54.55	32.11	264.82	41.19	6.865	15.31
Std	± 10.42	± 5.254		±1.72	±0.28	

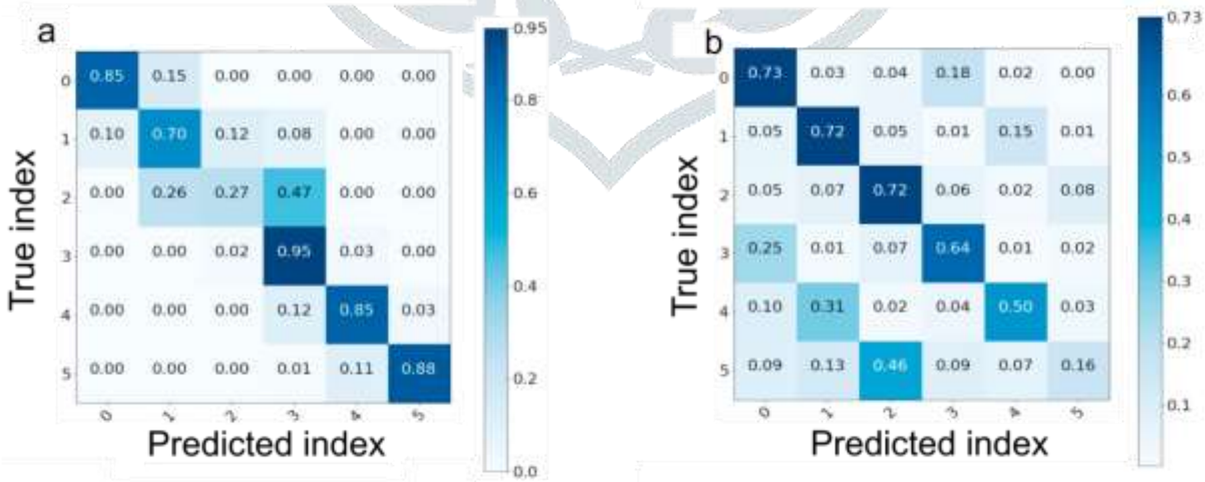

Fig. 3. Normalized confusion matrix for (a) ICRT and (b) VIT-MR datasets. Colour indicates the accuracy for each class as represented by the scalar number. Higher colour intensity in the diagonal of the matrix correspond to higher accuracy results of the model. Colour intensity are normalized with the maximum and minimum values for each dataset. Class indexes are referred in Table 2.

CONCLUSION

Table 3 findings show the TDNN network has an outstanding ability to adapt to task recognition that outperforms the other approaches considered. The TDNN's robustness is proven by the accuracy obtained on both datasets. The significant difference for VIT-MR dataset between micro and macro accuracy is due to phase distribution imbalances. This is seen as the difference varies from a minimum of 16 per cent for TDNN to a maximum of 22 per cent for SVM and thus assumes a significant value for each presented process. This statement is not true for ICRT dataset, as it provides more balanced distribution of phases. Most of the VIT-MR dataset's wrong recognitions occur between one step and its corresponding class of errors. The normalized confusion matrix shown in Figure 3 clearly reflects this fact. The difference in accuracy in phase recognition is also provided by the form of motion and sensor that more determines each phase. For robotic movements the best recognition output is obtained (Fig 3a last 3 labels) accompanied by human operator (Fig 3 first 3 labels of both datasets). The proposed method obtains low recognition efficiency when used for error detection, as shown by results from the last 3 VIT-MR labels in Fig3b. This implies that an APM-trained network is not suitable for recognizing errors that require more knowledge, perhaps of a higher level (such as those that are extracted from the environmental camera). As reported the calculation times of each method in Table 3 including the sum of training and prediction time for a single run. This measurement time is critical for evaluating the ability of various APM methods to be implemented on novel IMS plants with minimal effect on setup time and therefore costs.

REFERENCES

[1] Y. Goldberg, “A primer on neural network models for natural language processing,” *J. Artif. Intell. Res.*, 2016, doi: 10.1613/jair.4992.

[2] E. Smedler, S. Malmersjö, and P. Uhlén, “Network analysis of time-lapse microscopy recordings,” *Front. Neural Circuits*, 2014, doi: 10.3389/fncir.2014.00111.

[3] R. (200. R. from http://www.sciencedirect.com/science/article/pii/S0169555X12003819 Carley, J., Pasternack, G., Wyrick, J., & Barker, J. (2012). Significant decadal channel change 58–67years post-dam accounting for uncertainty in topographic change detection between *et al.*, “RESTORING THE FISH FAUNA CONNECTIVITY OF THE HÂRTIBACIU RIVER-RETIŞ DAM STUDY CASE (TRANSYLVANIA, ROMANIA).,” *Acta Oecologica*, 2017.

[4] H. Merchant, D. L. Harrington, and W. H. Meck, “Neural Basis of the Perception and Estimation of Time,” *Annu. Rev. Neurosci.*, 2013, doi: 10.1146/annurev-neuro-062012-170349.

[5] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. De Vries, M. J. N. L. Benders, and I. Isgum, “Automatic Segmentation of MR Brain Images with a Convolutional Neural Network,” *IEEE Trans. Med. Imaging*, 2016, doi: 10.1109/TMI.2016.2548501.

[6] D. Maturana and S. Scherer, “VoxNet: A 3D Convolutional Neural Network for real-time object recognition,” 2015, doi: 10.1109/IROS.2015.7353481.

[8] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, 2014.