2022 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT) | 978-1-6654-6090-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/BDCAT56447.2022.00028

Surrogate-based Digital Twin for Predictive Fault Modelling and Testing of Cyber Physical Systems

Hayatullahi Bolaji Adeyemo
School of Computer Science
University of Birmingham
Birmingham, UK
hba986@student.bham.ac.uk

Rami Bahsoon
School of Computer Science
University of Birmingham
Birmingham, UK
r.bahsoon@bham.ac.uk

Peter Tiňo
School of Computer Science
University of Birmingham
Birmingham, UK
p.tino@bham.ac.uk

***Abstract*—Cyber Physical Systems (CPS) pose a pressing need to ensure they are sufficiently reliable and continue to be dependable. It is, therefore, essential to test these systems to uncover any potential anomalies, which if not detected can lead to failure and/or cause loss or injury. Adequate or complete coverage of behaviours can be difficult to accomplish in CPS. We advocate a less expensive and easy-to-evaluate representation of the system via surrogate modelling. In this paper, we present a novel predictive fault modelling framework leveraging surrogate-based Digital Twin for probing for likely faults that can support software analysts and testers of CPS in their testing plans. The approach abstracts the CPS and uses a variant of Recurrent Neural Network known as Long Short-Term Memory (LSTM) surrogate model for forecasting. The forecasting can help in predicting multiple behaviours of the system components and the likely faults of systems under test; observations will consequently feed into the testing plans. Both direct and iterative (i.e. one-time and multiple-time varying steps) forecasting are supported as part of the framework. We evaluate our surrogate-based Digital Twins predictive modelling approach on two CPSs namely: water distribution system and air pollution detection system. The results show that our approach performed decently in predicting multiple time steps.**

I. Introduction

Cyber-Physical Systems (CPS) are characterised by a complex interaction between physical entities and computational processes with each entity inevitably contributing to the process of achieving the overall goal of the system [1]. Dependence on CPS has tremendously increased recently; therefore, there is a pressing need to ensure the systems are reliable enough [2] and continue to be dependable. Failure to identify and address anomalies or faults in the systems would likely cause injury, or at least discomfort to users and compromise the behaviour of the system. The complex interactions among physical components and computational processes of CPS has increased the challenge for planning test that can probe for likely behavioural faults. As a result, the need for providing cost effective and efficient means for predicting faults in CPS that can be fed into testing plans becomes more paramount.

Digital Twin (DT) technology, as it gradually creeps into software engineering activities, has the promise to provide a variety of benefits for fault modelling, prediction, analysis and testing for uncertainty. Researchers have recently discussed the concept of DT widely from the standpoint of simulation, characterising the twin as a collection of many models, data, simulations, and information [3] with a wide range of applications. DT is founded on the premise that the availability of cheap sensors, pervasive connectivity and the internet of things can hasten the capacity to capture data and feed it to digital versions of the physical systems under analysis. Industries with large scale products that have recorded a tremendous success with their adoption of DTs include automotive industry, healthcare services, manufacturing, power-generation, urban planning, etc.

Analysing processes and possible behaviours within the system can be expensive for complex systems, therefore, a more amenable representation of the process, popularly known as surrogate model, is usually employed. A surrogate model (also called emulator and meta-model) emulates the behaviour of a system by converting the computationally expensive optimisation problem into a less expensive one over a comprehensive space of the parameters involved [4].

The novel contribution of this paper is as follows: We define a novel predictive fault modelling and prediction framework that leverages surrogate-based Digital Twin with the objective of forecasting for likely faults attributed to varying range of typical or unforeseen behaviours and modes of interaction in CPS. The approach abstracts the CPS and uses Long Short Term Memory(LSTM) surrogate model for forecasting. The framework can support software analysts and testers in their testing plans of CPS behaviours. The specific contributions include: We contribute to concepts and foundation for predictive fault modelling in DTs. We report on the design of DT that leverages surrogate modelling as a novel approach. We describe multivariate multistep behaviour that can be predicted using our surrogate model-based DT. We use two CPS cases to evaluate our approach: water distribution system and air pollution detection system. The results show that our approach preformed decently in predicting multiple series.

The rest of this paper is structured as follows: We present related work and our approach in Section II and III respectively. The fault-based digital twin framework is described in Section IV. Section V reports on the evaluation and discusses the results. Section VI concludes with some directions to future work.

978-1-6654-6090-3/22/$31.00 ©2022 IEEE
DOI 10.1109/BDCAT56447.2022.00028

II. Closely Related Work

DT has been effective in designing, optimization, maintenance and monitoring of software systems for various concerns, such as performance, security, reliability, etc. Lee et al. [5] demonstrated the ability of a DT in examining various "what-if" situations.

DTs, for example, have been applied to the domain of CPS security [6] to detect anomalies. Eckhart and Ekelhart [7] created a knowledge-based intrusion detection system based on the idea that a CPS will display certain uncommon behaviour patterns during an assault. The authors provided rules that the system must follow under some normal conditions. Based on these rules, a basic DT that investigates rule breaches during runtime is created. The work was improved by employing a passive state replication method to imitate real-world systems using real-time data [8]. While emphasising that digital twin will provide new potential, Rubio et al. [9] presented a survey on anomaly detection [6] and evaluated the future trend.

III. Surrogate-based Digital Twin for Fault Prediction and Testing: A novel approach

This section presents a surrogate-based digital twin that is designed to predict varying and unforeseen behaviours of CPS. We present the novel formulation and description of the surrogate-based digital twins approach.

A. Surrogate Modelling Based Digital Twins

Digital twins often include simulation component symbiotically linked to the physical entities. This can be expensive to run repeatedly especially whenever engineers need to carry out optimisation, sensitivity analysis or risk analysis multiple times. However, application of DT as a surrogate model can drastically reduce the cost. This is because surrogate models basically approximates the simulation results and can then be used to replace the original computer simulation in activities like prediction, testing and optimisation. We envisioned that DTs will support prediction and testing plan identification of the system's behaviours.

B. Approach to building Surrogate-based Digital Twin

In CPS testing, the need for frequent evaluation of the system behaviour has led to the proposal of a more accurate and scalable approach to representing the behaviour of the system. This is demonstrated by surrogate models.

1) Problem Formulation: After collecting data from simulation and organising it into a supervised learning format, we also formulate the problem as a multivariate time series forecasting as follows:

$$\hat{y}_{j,t+1} = f(y_{j,t-k:t}, x_{j,t-k:t}, s_j) \tag{1}$$

where j is the index of each univariate time series within the multivariate series, $\hat{y}_{j,t+1}$ is the single-horizon forecast for the jth time series, while the parameters of the model's learnt function $f(.)$ are respectively the target observations ($y_{j,t-k:t} = \{y_{j,t-k}, ..., y_{j,t}\}$), external input to train the model ($x_{j,t-k:t} = \{x_{j,t-k}, ..., x_{j,t}\}$) and the static information (s_j) related to the system whose quantities are to be predicted (such as the initial conditions). The target observations and the external input are based on the size of the lookback window (k). In this formulation, we predict only one time-step ahead of each of the input signals.

In traditional time series prediction problem, the target is predicted for the next time frame based on the input features of the previous time frame. But in this work, given some input signals, the aim is to develop a model to accurately predict multiple time-step ahead of the multivariate signals, with a slight modification to Equation (1).

$$\hat{y}_{t+\tau} = f(y_{t-k:t}, x_{t-k:t}, U_{t-k:t+\tau}, s, \tau) \tag{2}$$

where $\tau \in \{1, 2, 3, ..., \tau_{max}\}$ is a discrete forecast horizon. U_t are the actual future inputs throughout the whole time horizon for the purpose of evaluation and x_t are historically observed inputs. Other notations remain the same as mentioned in Equation (1).

2) LSTM-based Multivariate Multi-Step Prediction in CPS: Like any other time-variant systems, CPS depict their inputs and outputs as signals. In this work, we model the behaviour of CPS using an LSTM, due to its modelling capability of signals as sequence data. The aim is to predict the behaviour of the system given some input signals. Another reason we selected LSTM is because of its capacity to address the challenges of vanishing and exploding gradients.

The LSTM has different structures (architectures) based on the nature of the problem (i.e. the number of input and output attributes). Considering the nature of the CPS, it is characterised by multiple inputs and outputs even though a particular output may be selected as a point of interest because the combination of the outputs defines the overall behaviour of the system. Therefore LSTM can be classified based on input-output feature. In other words, LSTM can be one-to-one, one-to-many, many-to-one and many-to-many. It becomes more challenging when the prediction is multivariate, which involves multiple quantities to be predicted at a time.

The prediction is done from two different strategies: iterative and direct approaches. Firstly, iterative approach basically produces multi-time forecasts by iteratively feeding samples to the model that forecasts a single output (as in Equation (1)). Multiple predictions are generated by repeating the process to create the trajectories. This approach is a generalisation of the one-step forecasting model to predict multiple steps. The drawback of the approach is the accumulation of the errors obtained at each time step through the entire forecasting horizon. On the other hand, direct approach is used to generate multi-step ahead prediction directly at a go. It is an adaptation of the single-step prediction to produce multi-step ahead forecast (as in Equation 2). The direct approach can mitigate the problem of error accumulation in the iterative approach. The encoder-decoder LSTM takes the inputs and maps them to a fixed-length vector through the encoder and predicts the output sequence by decoding the fixed-length vector through the decoder. The input size can be different from the output size and this has made them useful in a variety of problems

such as language translation, where the seminal work [10] on encoder-decoder focused on.

IV. Fault-based Digital Twin Prediction and Testing

Fault injection is the conventional method for empirically evaluating a system's dependability. An objective of injecting faults into the surrogate models is to ascertain that the model is a good representative of the system under test and can respond accurately to different fault types, similar to how the real system would respond to fault injection.

A. Framework of Digital Twin for predictive fault modelling and testing

The described framework aims to enhance testing, design optimisation, forecasting and predictive maintenance of a CPS after defining the use-case driven by DT by taking advantage of the time-indexed sensor data. Fig. 1 illustrates the proposed framework, which comprises the system modelling, data collection block, data processing block, the fault injection block, the time-series prediction block and the monitoring control block.

Data is collected by simulating the CPS under study while extracting important information that would be used to model the system. The data is cleansed before normalised in order to allow accurate training of the surrogate model. The framework presents a component for fault injection to introduce faults into the trained model and the simulator with the intent of carrying out fault-based testing. This can allow a successful carrying out of testing and optimisation of testing processes. This can also leverage DT capability to carry out forecasting the future behaviour of the twin when subjected to different operation conditions.

In this paper, we focus on the forecasting aspect of the framework, (the detail of which is presented in Section III-B), which can help in testing plans and predictive maintenance. decisions.

V. Results and Discussion

In this section, we instantiate our model on two CPS: air pollution monitoring and a water distribution systems. The purpose is to exemplify the approach and to show how it can be utilised in prediction of CPS behaviour. We describe the two systems whose datasets were collected to create their surrogate-based DTs as a way to exemplify and evaluate our approach.

A. Description of datasets

The first experimental time series dataset [11] is the result of the emulation of water flow between tanks and pumps, valves, flow and pressure sensors. The dataset contains 2420 samples with 42 features including information of 8 tanks, 6 pumps, 4 flow sensors, 22 valves. The second dataset [12] is on air quality and describes hourly updates on the pollution and weather levels in Beijing, China from January 2010 up to December 2014. The dataset contains 43800 samples covering the span of five years with an hourly frequency.

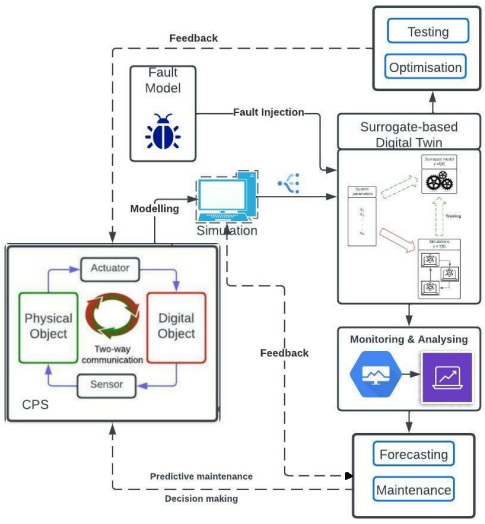

Fig. 1: Framework of the Digital Twin

B. Evaluation Metrics

There are numerous metrics to evaluate the usefulness of a learning model. We choose root mean squared error (RMSE), normalised root mean squared error (NRMSE) and coefficient of determination (R-squared) to measure the model's performance. The metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - p_i)^2} \tag{3}$$

$$NRMSE = \sqrt{\frac{1}{n}\frac{\sum_{i=1}^{n}(a_i - p_i)^2}{max(a_i) - min(a_i)}} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(a_i - p_i)^2}{\sum_{i=1}^{n}(a_i - \bar{a})^2} \tag{5}$$

where n, a_i and p_i respectively represent the number of time samples, actual and predicted values of the ith sample with $\bar{a}$ as the average (mean) value of the actual observations. The $max(a_i)$ and $min(a_i)$ denote the maximum and minimum of the actual observations respectively. RMSE can be normalised using mean, range, standard deviation or interquartile range; but we normalised using the range in order to give better interpretable values that are between 0 and 1. The normalised rmse value that is closer to 0 is considered better than the one closer to 1. This makes it easy to interpret the values.

C. Discussion of results

We report on the prediction of behaviours of water distribution system (comprising 3 tanks and 1 pump) and an air pollution monitoring system. We applied two different techniques to forecast the future trajectory namely iterative and direct methods. The errors in iterative method keep increasing due to the accumulation of prediction error at each timestep.

TABLE I: Prediction errors for different epochs for Water distribution and Air pollution systems

Epochs	Metrics	Tank1	Tank2	Tank3	Pump1	Pollution	Dew point	Temperature	Pressure	Wind speed
50 epochs	RMSE	167.55	225.42	229.36	0.1152	5.15	3.40	0.96	0.52	9.08
	Normalised rmse	0.2558	0.3436	0.2433	0.2513	0.5725	0.3094	0.3196	0.2607	0.2257
	R-squared	0.9286	0.8815	0.9711	0.9027	0.4704	0.5834	0.9473	0.6643	0.4665
100 epochs	RMSE	184.97	146.11	256.55	0.0960	15.99	2.27	3.39	2.42	5.21
	Normalised rmse	0.2824	0.2227	0.2722	0.2094	0.4543	0.1511	0.3082	0.2205	0.2926
	R-squared	0.9129	0.9502	0.9639	0.9324	0.5611	0.6360	0.2822	0.2694	0.1217
200 epochs	RMSE	216.22	227.13	347.20	0.0675	53.08	0.63	0.97	1.03	5.09
	Normalised rmse	0.3301	0.3462	0.3684	0.1473	0.4423	0.2097	0.1932	0.2579	0.6688
	R-squared	0.8810	0.8796	0.9339	0.9667	0.3993	0.6652	0.5586	0.1751	0.7167

Therefore, we based our evaluation on the direct method. We implemented an LSTM with 100 neurons each for encoder and decoder, Adam optimiser with learning rate of 0.001, and two layers each for the encoder and the decoder. Table I shows the results of the prediction under different values of epochs while evaluating with the metrics RMSE, normalised rmse, and R-squared (R^2) values. Based on the R^2 values obtained, we observe that the prediction errors are quiet similar, showing 100 epochs having the overall lowest average error for the water distribution system. For each predicted feature, there is a RMSE value, which makes it different from most of the work in the literature, where the focus is only on predicting single feature at a time. Even though the RMSE values are high, the prediction seems to be promising due to the relatively high R^2 values.

With this approach, each sensory component of the monitoring system can be predicted reliably ahead of time and the injection of faults in the sensory data could lead to high deviation in the trajectory of the behaviour after the forecasting. This will be evident in the sense that each of the components are connected in one way or the other. This would also help in understanding fault propagation in the system.

VI. Conclusion and future work

We have presented a surrogate-based DT model for fault modelling and prediction of CPS behaviours that can feed into testing plans of CPS for unforeseen behaviours under uncertainties. The approach makes a novel use of surrogate-based prediction of multivariate multi-steps to predict the behaviour of CPS and its constituent components over time, under different operational scenarios. The prediction technique was implemented using a variant of Recurrent Neural Network known as Long Short Term Memory (LSTM), where we investigated the predictive accuracy of direct and iterative methods, while predicting multiple behaviours at the same time. We have formulated a fault model for fault injection in the Digital Twin. The proposed work can assist CPS testers and engineers in predictive maintenance and in generating testing plans for likely faults and unforeseen modes of interaction. The digital twin can emulate potentially "rare" faults that can reveal unexpected behaviours of CPS. The surrogate model-based DT can also compress time to reveal the propagation of faults over an extended time. These faults can be linked to typical extreme case scenarios that can break or halt the system. The insights gained can help CPS designers and architects to refine the architecture and design of these systems, in an attempt to safeguard the system against these faults. The effectiveness of the refinements can then be judged based on the system's ability to converge towards acceptable behaviour.

Our future work will extend the framework to support automated online testing leveraging DT capabilities and will leverage MAPE-K primitives of autonomous computing to dynamically update the DT following changes in the physical system.

References

[1] Martin W Hoffmann, Somayeh Malakuti, Sten Grüner, Soeren Finster, Jörg Gebhardt, Ruomu Tan, Thorsten Schindler, and Thomas Gamer. Developing industrial cps: A multi-disciplinary challenge. *Sensors*, 21(6):1991, 2021.

[2] Amit Kumar Tyagi and N Sreenath. Cyber physical systems: Analyses, challenges and possible solutions. *Internet of Things and Cyber-Physical Systems*, 2021.

[3] Michael Schluse and Juergen Rossmann. From simulation to experimentable digital twins: Simulation-based development and operation of complex technical systems. In *2016 IEEE International Symposium on Systems Engineering (ISSE)*, pages 1–6. IEEE, 2016.

[4] Mohammed Reza Kianifar and Felician Campean. Performance evaluation of metamodelling methods for engineering problems: towards a practitioner guide. *Structural and Multidisciplinary Optimization*, 61(1):159–186, 2020.

[5] Dongmin Lee, SangHyun Lee, Neda Masoud, MS Krishnan, and Victor C Li. Digital twin-driven deep reinforcement learning for adaptive task allocation in robotic construction. *Advanced Engineering Informatics*, 53:101710, 2022.

[6] Qinghua Xu, Shaukat Ali, and Tao Yue. Digital twin-based anomaly detection in cyber-physical systems. In *2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST)*, pages 205–216. IEEE, 2021.

[7] Matthias Eckhart and Andreas Ekelhart. Securing cyber-physical systems through digital twins. *Ercim News*, 2018(115), 2018.

[8] Matthias Eckhart and Andreas Ekelhart. Towards security-aware virtual environments for digital twins. In *Proceedings of the 4th ACM workshop on cyber-physical system security*, pages 61–72, 2018.

[9] Juan Enrique Rubio, Cristina Alcaraz, Rodrigo Roman, and Javier Lopez. Analysis of intrusion detection systems in industrial ecosystems. In *SECRYPT*, pages 116–128, 2017.

[10] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.

[11] Luca Faramondi, Francesco Flammini, Simone Guarino, and Roberto Setola. A hardware-in-the-loop water distribution testbed dataset for cyber-physical security testing. *IEEE Access*, 9:122385–122396, 2021.

[12] Saverio De Vito, Ettore Massera, Marco Piga, Luca Martinotto, and Girolamo Di Francia. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2):750–757, 2008.