ISSN No:-2456-2165

A Review on Various Techniques for Data Analysis in Cyber-Physical Systems

T. Rajalakshmi
Ph.D., Research Scholar
Department of Computer Science
Erode Arts & Science College
Erode, Tamil Nadu, India

S. Sathappan
Associate Professor,
Department of Computer Science
Erode Arts & Science College
Erode, Tamil Nadu, India

Abstract:- Cyber-Physical System (CPS) is interconnected physical systems for mission-critical processes. The emerging CPS should be more robust, coordinated, distributed and linked in terms of capacity, versatility, power, protection and usability. The huge volume of data in CPS can be handled by big data technologies. The CPS data come from many sources in many formats, it is difficult to analyze and extract the useful information. This problem can be solved by using deep learning with advanced data processing techniques. It helps to obtain information from wide and commonly multi-source data sets capable of not only containing structured but also unstructured and semi-structured information. There are various techniques has been developed to analysis the huge volume of CPS data for knowledge extraction or decision making process. This paper presents a survey on various techniques for data analysis in CPS. In this paper different techniques for data analysis in CPS developed by previous research are studied in detail. The limitation of each techniques and provide a suggestion for further improvement in CPS data analysis.

Keywords:- *CPS;Cyper Physical System ; Machine Learning, Data Analiss, Long Short Term Memory,Deep Learning.*

I. INTRODUCTION

Cyber-Physical Systems (CPSs) is a recent system that involves different interconnected elements such as sensors, actuators and controllers that need to instantly share, analyze and integrate heterogeneous data. The main functionalities of CPS include forecasting, synchronizing and controlling the sensing information in the real-world scenario. It creates two main problems for the development of CPS: the quantity of available information from different data sources to be processed at any given time and the choice of process controls as a result of the information obtained.

To maintain effective control of the underlying physical processes, an optimum balance must be achieved between data available and its quality [2]. The characteristic of CPS are distributed management or control, high level automation, real-time requirements in performance, dynamic re-organisation or reconfiguration, multi-scaling system control features etc. Mostly, wide range of CPS applications is used in the modern world such as electric power grid and energy systems, etc. The extraction of useful knowledge from CPS data is more difficult because of its heterogeneity property.

Machine Learning (ML) techniques have achieved an impressive result to solve this problem. They can reduce development cost and provide practical solution to make decision making and control problems in CPS data. Deep learning [3] is a machine learning that utilizes different layers of nonlinear processing units to enhance the processing results of CPS data. The main intention of this article is studying in detailed information on various techniques for data analysis CPS. In addition, their limitations are also addressed to suggest further improvement on data analysis in CPS.

II. A REVIEW ON VARIOUS TECHNIQUES FOR CYBER-PHYSICAL SYSTEMS DATA ANALYSIS

Pasqualetti et al. [4] proposed a centralized and distributed monitors for detect and identify the attacks. Initially, Optimal centralized attacks detection and identification monitors were designed. The waveform relaxation method was used to build an optimal distributed attack detection filter. Also, a sub-optimal detection of distributed attacks was designed to ensure performance guarantees. However, this method has highly computationally complex.

Spezzano et al [5] proposed an algorithm, for pattern detection in CPS. It detected Spatio-temporal patterns in a network of heterogeneous real-time data sources. The local stochastic multi-agent search strategy was allowed to perform independently and interact with immediate neighbours in asynchronous way. This algorithm was scalable through large data sets by synchronism. However, this algorithm cannot detect patterns in CPS very efficiently.

Ntalampiras et al [6] proposed an Integrity Attack Diagnosis Systems (IADS) method for automatic identification of integrity attacks in CPS**.** In this novel method, a feature set was designed that captured the features of each attack in the wavelet and spectral domains when its allocation was learned by pattern recognition algorithms. In addition, a new detection aspect was introduced to deal with previously unseen attack forms.

However, this method was highly time-consuming to build the decision pattern for attack identification in CPS.

Shin et al. [7] introduced Recurrent Neural Networks (RNN) technique for detection and identification of sensor attacks in CPS. Initially, data obtained from the sensor node was normalized at pre-processing phase. Then, normalized data was used in the learning phase where the weights of the model were optimized with the training data. Based on the learning model, the RNN classified the sensors into seven states at the prediction phase. However, training of RNN model was more difficult.

Goh et al. [8] proposed a Recurrent Neural Network (RNN) approach for behavioural-based intrusion detection in CPS. Initially in this approach, a Long Short Term Memory-RNN (LSTM-RNN) was designed as a predictor to find the normal behavior. Concurrently, a cumulative method was used to find the abnormal behavior in the network. Based on the abnormal behavior in the network, intrusions were detected. However, it cannot validate false positive.

Cui et al. [9] proposed an Oriented Cuckoo Search algorithm to improve the Distance Vector-Hop (DV-Hop) method during the classification of CPS data. In this algorithm, the combination of two different random distributions, Cauchy distribution and Levy distribution, was managed by the global search capacity in the OCS algorithm. The DV-Hop method was also implemented to improve the accuracy of the algorithm. However, this algorithm was high computational complexity problem.

Zhao et al. [10] proposed a parameter-free Incremental Co Clustering (PFICC) algorithm that deals with multi-modal data in dynamic manner. Initially, single modality similarity measure was extended to multiple modalities. Also three processes were performed such as cluster generation, cluster fusion, and data partition. These processes were used for integrating new arriving objects and dynamically adjusting clustering pattern. Furthermore, the importance of feature modalities were measured using an adaptive method on the basis of intra-cluster scatters. However, computational complexity is high due to high dimensionality multi-modal data.

Ghazi et al. [11] proposed a method for anomaly behaviour and miuse detection in the network. This method was suitable for Supervisory Control and Data Acquisition (SCADA) scheme at the highest level of CPS. Here, the Neural First Order Hybrid Petri Net Model (NFOHPN) with online fast Independent Component Analysis (ICA) model was used for anomaly detection. The independent components of the sampled data were extracted using ICA. The extracted features were used in NFOHPN to detect the cyber-attacks in CPS. However, this method is computationally expensive.

Yuan et al. [12] proposed an Embedding Clustering Based Deep Hypergraph (ECDHG) model for online review analysis in CPS. In this model, external knowledge was introduced by applying the pre-training word embedding for express reviews. After that, Hierarchical Fast Clustering Algorithm (H-CFS) was used for identifying semantic units under the control of semantic cliques. Moreover, the high order textual and semantic features of reviews were extracted by using convolutional neural networks. Based on the extracted high order features, the hyper graph was constructed for classifying sentiment reviews.

Amuthadevi et al. [13] proposed a fuzzy logic concept based CPS to detect the air pollution in urban areas. Fuzzy logic was used for detection of the extremely air polluted industry and region. Initially, the required data was routed between nodes using the routing protocols to monitor and control. After that, the data was processed on the basis of fuzzy logic concept to observe and validate the air control. However, the selection of proper membership function was more difficult in this method.

Ruan et al. [14] proposed a Genetic Algorithm-Support Vector Machine (GA-SVM) predictor for agricultural CPS. The agricultural CPS data was split into number of granules using Quartile-Median granulation, Fuzzy granulation and Min-Median-Max granulation methods. The number of granules was processed in GA-SVM predictor where GA was used to fine tune the parameters of SVM and SVM was used to predict the agricultural yield.

Kang & Chung et al. [15] developed a Novel deep learning called DeepRT which focus on supporting predictable temporal and spatial inference performance when deep learning models are used under unpredictable environments. Also, DeepRT determined the proper level of compression of deep learning models at runtime according to the applications of CPS data. It supported the timeliness of inference tasks in a robust and energy-efficient way. The DeepRT model was work well in the features with sufficient training data, but they could not extrapolate beyond the training space. It is known that a feature space that lacks training data generally has much error rate. Hence, it is necessary to identify the training space and avoid extrapolating beyond the training space. Moreover, there is no guarantee that the training data is even representative of accurate learning.

Table 1 illustrates an overview of merits and demerits of above discussed various techniques for CPS data analysis.

Ref. No.	Methods Used	Merits	Demerits
[4]	Waveform relaxation method	Robust to system noise.	High computational complexity.
[5]	Machine learning algorithm	Highly scalable to very large datasets.	Detection pattern in CPS is not efficient.
[6]	IADS	Provides almost perfect identification of integrity attacks.	The decision pattern for attack detection in the CPS is consumes very high time.
[7]	RNN	Guarantees to identify deceptions as cyber-physical attacks.	Does not operate under complex driving conditions such as stop, slope, etc.
[8]	LSTM-RNN	Low false positive rate.	In LSTM-RNN model, the validation of false positive rate does not occurs.
[9]	OCS	High precision.	Easily falls into the local optimal solution and the slow rate of convergence
[10]	PFICC	Effectively deal the multi-modal data in CPS.	For high dimensionality multi model data, the computational complexity is high.
[11]	SCADA	The convergence speed of the SCADA makes it particularly useful to detect critical CPS intrusions.	The method is complicated in terms of trained techniques, analysis and programmers.
[12]	H-CFS	Can effectively assist the sentiment classification model with detecting meaningful latent semantic units and characterizing the original text accurately.	The hyper edge size is used in H-CFS which is obtained by a large number of experiments and it may not be optimal for some datasets.
[13]	Fuzzy logic concept	High packet delivery ratio.	The selection of proper membership function was more difficult in this method.
[14]	GA-SVM	Reduce computational time with equivalent accuracy.	Does not handle more complicated data.
[15]	DeepRT	Minimize the loss of inference accuracy.	Performance of DeepRT is degraded due to Feature space lacking

Table 1:- Comparison of various techniques for CPS data analysis

III. CONCLUSION

In this article, various techniques are reviewed to analysis the CPS data based on their merits and demerits. On the basis of review between different techniques, it is clear that DeepRT can effectively minimize the loss of inference accuracy during the CPS data analysis.In future, this problem will be focused and develops an efficient technique for CPS data analysis.

REFERENCES

[1]. Park, K. J., Zheng, R., & Liu, X. (2012). Cyber-physical systems: Milestones and research challenges
[2]. Petrovski, A., Rattadilok, P., & Petrovski, S. (2015, September). Designing a context-aware cyber physical system for detecting security threats in motor vehicles. In Proceedings of the 8th International Conference on Security of Information and Networks (pp. 267-270)
[3]. Konstantinou, C., Maniatakos, M., Saqib, F., Hu, S., Plusquellic, J., & Jin, Y. (2015, May). Cyber-physical systems: A security perspective. In *2015 20th IEEE European Test Symposium (ETS)* (pp. 1-8). IEEE
[4]. Pasqualetti, F., Dörfler, F., & Bullo, F. (2013). Attack detection and identification in cyber-physical systems. IEEE transactions on automatic control, 58(11), 2715-2729.
[5]. Spezzano, G., & Vinci, A. (2015). Pattern detection in cyber-physical systems. Procedia Computer Science, 52, 1016-1021.
[6]. Ntalampiras, S. (2016). Automatic identification of integrity attacks in cyber-physical systems. Expert Systems with Applications, 58, 164-173.
[7]. Shin, J., Baek, Y., Eun, Y., & Son, S. H. (2017, November). Intelligent sensor attack detection and identification for automotive cyber-physical systems. In 2017 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1-8). IEEE.
[8]. Goh, J., Adepu, S., Tan, M., & Lee, Z. S. (2017, January). Anomaly detection in cyber physical systems using recurrent neural networks. In 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE) (pp. 140-145). IEEE.

[9]. Cui, Z., Sun, B., Wang, G., Xue, Y., & Chen, J. (2017). A novel oriented cuckoo search algorithm to improve DV-Hop performance for cyber–physical systems. Journal of Parallel and Distributed Computing, 103, 42-52.
[10]. Zhao, L., Chen, Z., & Yang, Y. (2017). Parameter-free incremental co-clustering for multi-modal data in cyber-physical-social systems. IEEE Access, 5, 21852-21861.
[11]. Ghazi, Z., & Doustmohammadi, A. (2018). Intrusion detection in cyber-physical systems based on petri net. Information Technology and Control, 47(2), 220-235.
[12]. Yuan, X., Sun, M., Chen, Z., Gao, J., & Li, P. (2018). Semantic clustering-based deep hypergraph model for online reviews semantic classification in cyber-physical-social systems. IEEE Access, 6, 17942-17951.
[13]. Amuthadevi, C., Priya, J. S., & Madhusudhanan, B. (2019). Validation of multicast routing in cyber physical systems monitoring air quality. Cluster Computing, 22(2), 3917-3923.
[14]. Ruan, J., Jiang, H., Li, X., Shi, Y., Chan, F. T., & Rao, W. (2019). A Granular GA-SVM Predictor for Big Data in Agricultural Cyber-Physical Systems. IEEE Transactions on Industrial Informatics.
[15]. Kang, W., & Chung, J. (2019). DeepRT: predictable deep learning inference for cyber-physical systems. Real-Time Systems, 55(1), 106-135