A Privacy Preserving Federated Learning (PPFL) Based Cognitive Digital Twin (CDT) Framework for Smart Cities

Sukanya Mandal
Dublin City University, Dublin 9, Ireland
sukanya.mandal2@mail.dcu.ie, sukanyam.research@gmail.com

Abstract

A Smart City is one that makes better use of city data to make our communities better places to live. Typically, this has 3 components: sensing (data collection), analysis and actuation. Privacy, particularly as it relates to citizen's data, is a cross-cutting theme. A Digital Twin (DT) is a virtual replica of a real-world physical entity. Cognitive Digital Twins (CDT) are DTs enhanced with cognitive AI capabilities. Both DTs and CDTs have seen adoption in the manufacturing and industrial sectors however cities are slow to adopt these because of privacy concerns. This work attempts to address these concerns by proposing a Privacy Preserving Federated Learning (PPFL) based Cognitive Digital Twin framework for Smart Cities.

Introduction

Cities are becoming smarter with the integration of IoT (Internet of Things) devices, AI, and Digital Twin (DT) technologies with the main aim being to improve the quality of life and well-being of citizens, promote urban sustainability, and ensure inclusive socioeconomic growth. Technologies such as IoT analytics and DT visualizations could provide better insights by adding a layer of cognition to these existing technologies. This could also enable better information sharing, better decision making as well as better citizen engagement. To be able to leverage these cognitive capabilities, it is important to utilize advanced AI capabilities requiring massive amounts of data. This raises the question of data privacy as Smart Cities (SC) data typically encompasses lot of vital personal information. To address this challenge, in my thesis I am investigating the use of federated intelligence as an addition to the cognitive layer.

The introduction of a Cognitive Digital Twin (CDT) (Zheng, Lu, and Kiritsis 2022), with SC data has the potential to result in superior city performance in terms of operational efficiency, predictive abilities, autonomous decision making, and citizen engagement compared to traditional DT models. To realize a CDT, effective utilization of massive amounts of data is required, but data privacy concerns can slow down adoption. This research investigates using a Privacy Preserving Federated Learning (PPFL) based CDT framework to address data privacy concerns more effectively than traditional security methodologies.

This leads to the following research hypotheses underpinning this thesis:

- A CDT provides superior predictive capabilities, allowing SC to anticipate future scenarios more accurately compared to traditional DTs.
- A CDT supports more efficient autonomous decision-making, leading to improved city operations compared to traditional DTs.
- A CDT facilitates increased citizen engagement through enhanced interactive experiences and visualizations compared to traditional DTs.
- PPFL based CDT addresses data privacy concerns in SC more effectively by reducing the need for centralized data storage and enabling more decentralized processing of sensitive data.

These hypotheses posit that a PPFL based CDT framework for SC not only delivers superior performance in city operations but also offer better solutions to address data privacy concerns inherent in SC frameworks.

Background & Related Work

A DT is a digital replica of a real-world physical environment that simulates the attributes and behaviors of that system, for the purpose of enabling measurements, simulation and experimentations with the digital version to better understand the physical counterpart. A Hybrid Digital Twin (HDT) is an extension of DT in which siloed DT models are intertwined to predict behaviors of the physical counterpart before it happens. A CDT is an extension of HDT that incorporates cognitive features that enables

sensing complex and unforeseen patterns and reasoning about dynamic strategies for optimizations, leading to a continuously evolving
system that updates its digital version and behavior.

The core idea of Federated Learning (FL) is to train machine learning models on datasets that are distributed across different devices or parties, which can preserve the local data privacy. PPFL focuses on the privacy preserving mechanism of FL by employing techniques such as Homomorphic Encryption (HE), Secure Multi-party Computation (SMC) and Differential Privacy (DP).

In the recent years there has been much research in CDTs but has been mostly focused on the manufacturing and process industry. The COGNITWIN project (Abburu et al. 2020) proposes a CDT framework focused on the process industry. The CITYPULSE project (Puiu et al. 2016) is more related to my work as it proposed a SC framework utilizing knowledge graph and semantics and mainly covers (i) data acquisition (ii) semantic interoperability, (iii) real-time data analysis and event detection, and (iv) SC application development support. Finally, the survey presented in (Ramu et al. 2022) discusses the potential of using FL enabled DT for SC. These studies, along with many others, motivated me to pursue the presently proposed research direction.

Methodology & Work to Date

The high-level approach I propose to integrate FL with CDT is illustrated in Figure 1. In FL, the privacy of data is preserved at the data source transmitting only the updates after local training in the edge device and therefore the raw sensitive information is not required to be communicated over the network.

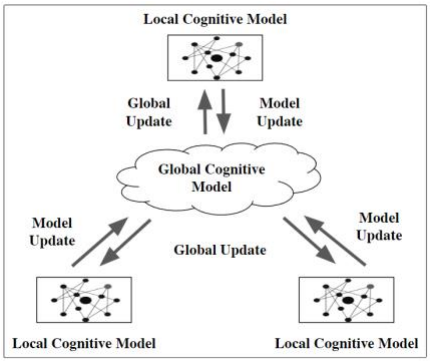

Figure 1: Integrating FL with CDT.

To date I have developed a reference architecture for a PPFL based CDT framework for SC and its various components as illustrated in Figure 2. The Cognitive Layer and the Federated Intelligence layers are where novel research contributions are targeted for the remainder of my PhD.

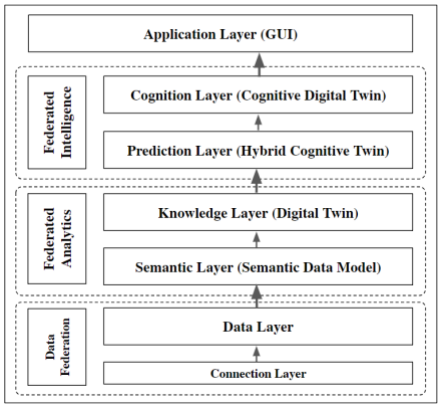

Figure 2: Reference architecture of PPFL based CDT framework for Smart Cities.

Future Work

My research roadmap to the end of my PhD is as follows:

- Ontology and Knowledge Engineering: The data model layer for smart cities building on existing ontologies. (in progress)
- Create a CDT Framework using Ontologies and Knowledge Graph capabilities capable of knowledge representation and reasoning from data feeds of the previously mentioned ontology. (in progress)
- Create a GNN (or other relevant methodology) based framework to form the cognition capability of the CDT framework for advanced knowledge representation and reasoning capabilities. (to do) [Novel contribution]
- Design, implement and test - PPFL methodologies particularly for CDT to incorporate with the CDT framework mentioned in the above step. (to do) [Novel Contribution]

References

Xiaochen Zheng, Jinzhi Lu & Dimitris Kiritsis 2022. The emergence of cognitive digital twin: vision, challenges and opportunities, International Journal of Production Research, 60:24, 7610-7632, DOI: 10.1080/00207543.2021.2014591.

Abburu, S.; Berre, A.; Jacoby, M.; Roman, D.; and Stojanovic, L. 2020. COGNITWIN – Hybrid and Cognitive Digital Twins for the Process Industry. IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), Cardiff, UK, 2020, pp. 1-8, doi: 10.1109/ICE/ITMC49519.2020.9198403.

Puiu, D. et al., "CityPulse: Large Scale Data Analytics Framework for Smart Cities," in IEEE Access, vol. 4, pp. 1086-1108, 2016, doi: 10.1109/ACCESS.2016.2541999.

Ramu, S.; Boopalan, P.; Pham, Q.; Maddikunta, P.; Huynh-The, T.; Alazab, M.; Nguyen, T.; and Gadekallu, T. 2022. Federated learning enabled digital twins for smart cities: Concepts, recent advances, and future directions. Sustainable Cities and Society, 79, p.103663, 2022, https://doi.org/10.1016/j.scs.2021.103663.