MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230961

Digital Twins for More Precise and Personalized Treatment

Nilmini WICKRAMASINGHE[a,1], Nalika ULAPANE[a], Elliot B SLOANE[b] and VijayGEHLOT[b]
[a] *Swinburne University of Technology, Australia*
[b] *Villanova University, USA*
ORCiD ID: Nilmini Wickramasinghe https://orcid.org/0000-0002-1314-8843

Abstract. The use of Digital Twins (DTs) or the digital replicas of physical entities has provided benefits to several industry sectors, most notably manufacturing. To date, the application of DTs in the healthcare sector has been minimal, however. But, as pressure increases for more precise and personalized treatments, it behooves us to investigate the potential for DTs in the healthcare context. As a proof-of-concept demonstration prior to working with real patients, we attempt in this paper, to explore the potential for creating and using DTs. We do this in a synthetic environment at this stage, making use of data that is all computer-generated. DTs of synthetic present patients are created making use of data of synthetic past patients. In the real world, the clinical objective for creating such DTs of real patients would be to enable enhanced real-time clinical decision support to enable more precise and personalized care. The objective of the numerical experiment reported in this paper, is to envisage the possibilities and challenges of such an approach. We attempt to better understand the strengths and weaknesses of applying DTs in the healthcare context to support more precise and personalized treatments.

Keywords. clinical decision making, clinical decision support, digital twin, personalized medicine, time series modelling

1. Introduction

A Digital Twin (DT) is a digital replica of a whole or a part of a physical-world entity. These twins connect with the physical entity through some form of data transfer and can simulate the physical entity's characteristics [1]. There is some debate over the distinction between a DT, and for instance, the well-known concept of 'model' that often comes in 'model-based control' in engineering. An interpretation that helps in this debate is to view a 'model' as a standalone entity, and to view a DT as a 'model' that is connected through data linkages. The DT will be participating in some process control in real time. This concept of DT has over time revolutionized certain practices such as manufacturing, product design and process control.

DTs have more recently found few applications in the healthcare sector as well (e.g., genomics, aged care, dementia, cancer care) [1], [2]. The objective in healthcare has been to construct precise models of patient progression, to compute more personalized interventions. There are several features that make DTs attractive to healthcare. Some features can be identified as convenient data visualization, and enabling of reasoning,

[1] Corresponding Author: Nilmini Wickramasinghe, email: nilmini.work@gmail.com or N.Wickramasinghe@latrobe.edu.au

experimenting, synthesizing, and forecasting. These can be done more efficiently, cost effectively, and risk aversely through DTs, than meddling with a physical entity. Realizing such usefulness of DTs, we present in this paper, a way to formulate DTs of patients, making use of the patient's own data where possible, and data of other patients where required.

The clinical objective behind constructing such DTs would be to enable learning patterns from a patient's historical data. This would enable optimizing present interventions by studying forecasts of patient progression through the DT.

Our paper serves as a proof-of-concept demonstration prior to making arrangements to work with real patients. Therefore, the work in this paper is carried out in a synthetic environment making use of data that is all computer-generated. By reflecting on the experience from this numerical experiment, we report in this paper some possibilities and challenges that can be envisaged to arise when going forward to do DTs with real patients and real data.

2. Methods

In this section, we present a synthetic data-based numerical experiment we conducted to understand the possibility of constructing DTs of patients.

Suppose we have a diabetic patient whose blood glucose level (denoted by the scaled variable y_n) at time instance n, varies according to (1) in which the scaled variable x_{n-1} denotes the insulin dosage taken at time instance $n-1$. The terms $\alpha^{(k)}$, $\beta^{(k)} \in \mathbb{R}$ are coefficients. The term $\epsilon_{(n)}$ indicates noise at time instance n.

$$y_n = \alpha^{(k)} y_{n-1} + \beta^{(k)}(x_{n-1} - 0.5) + \epsilon_{(n)} \quad (1)$$

We generate Ten different governing equations by varying $k = 1, 2, \ldots, 10$. These Ten equations represent Ten patients in this example. The governing equations differ by their coefficients (i.e., $\alpha^{(k)}$ and $\beta^{(k)}$). Coefficients were generated under the following rule: $\alpha^{(k)} = 1$ for all k and $\beta^{(k)}$ for any k can be a random number such that $0 < \beta^{(k)} < 1$. The generated coefficients and the corresponding equations are provided in the Results section in Appendix A provided as reference [3]. The rationale behind selecting the structure in (1) is that it is intuitive and easy to interpret while it ensures reachability for y_n to any $y_{(ref)} \in \mathbb{R}$ through iteration. It must be noted that these original equations that govern patients will remain unknown to us. Only the input and output data will be available. Therefore, our objective is to discover DTs that mimic the behavior of these equations.

Now suppose we observe these Ten patients over 100 equally spaced time instances, indicated by $n = 1, 2, \ldots, 100$. The blood glucose level of each patient is monitored and a suitable insulin dosage (as determined by a medical professional) is given at each time instance. Then, the response of blood sugar of each patient at the subsequent time instance is monitored.

For simulation, we generated a unique trajectory of insulin dosage using smooth splines for each patient, such that the constraint $0 \leq x_n \leq 1$ is satisfied for all n. An example trajectory is shown in Figure 1(a). The generated trajectories for all Ten patients are plotted in Appendix A [3], Results section. The implication of the constraint $0 \leq x_n \leq 1$ is as follows. x_n in our case is a scaled variable denoting insulin dosage. Therefore, $x_n = 0$ and $x_n = 1$ indicate minimum and maximum allowable dosages, respectively. Then, by using these insulin trajectories, the corresponding blood glucose

trajectory for each patient was calculated using the corresponding governing equation. The initial conditions (i.e., y_1 and x_1) for any patient could take a random number between 0 and 1. The noise constraint at each iteration was set such that the absolute value of noise (i.e., $|\epsilon_{(n)}|$) does not exceed 5% of the absolute value of the blood sugar level at the previous iteration (i.e., $|y_{n-1}|$). An example trajectory of blood glucose is shown by the blue circles in Figure 1(b). The blood glucose trajectories of all Ten patients are plotted in Appendix A [3], Results section.

We herein carry out two tests. The first one is named the "large n case" and the second one is named the "small n case". The large n case refers to the situation where a patient has a large amount of data; more specifically, the patient has data spanning across a time such that $n \geq \gamma$ where γ is some real-valued threshold. Thus, the small n case means a situation where a patient has little data, i.e., spanning across a time such that $n < \gamma$. For our analysis in this paper, we set $\gamma = 100$.

For the large n case we assume that a DT can be discovered from the patient's own data. But for the small n case we assume that the patient's own data is insufficient to discover an adequate DT. Therefore, we investigate for the small n case, whether DTs can be discovered with the aid of other DTs that have been discovered from patients who have satisfied the large n constraint. Thus, we propose estimating DTs in two ways as described in equations (2) and (3); the former is for the larger n case and the latter is for the small n case. Equations (2) and (3) are defined for the two-variable case, i.e., one independent variable and one dependent variable to correlate with the diabetes example we consider. a_i, b_i, d_i, c and e in (2) and (3) are real-valued coefficients. $DT^{(k)}$ in (3) is the DT of the k^{th} patient who has satisfied the large n constraint. The two tests conducted are described below.

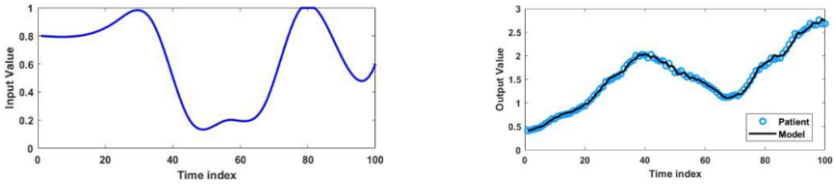
Test 1: The large n case. Using the available data of patient Number 8, it was attempted to estimate a DT from the patient's own data, and maintain the patient's output response y_n at $y_n = 1.5$ within $n = 101$ to 200 , by calculating x_n by solving the minimum norm minimization problem specified in (9) in Appendix A [3] for each iteration of n—Note: for interested readers we have provided a generic formulation of our DT inclusive of higher dimensions in Appendix A [3]. Results from this test are plotted in Figure 1(b) and Figure 2.

Test 2: The small n case. We assume that only the first 20 instances of the data of patient Number 8 is available to us. Thus, in this case patient Number 8 falls within the small n case. Then, the remaining Nine of the Ten patients generated were considered as past patients that satisfy the large n constraint. Thus, for those Nine patients, DTs were estimated from their own data following equation (2). Then those DTs were used to approximate a DT for patient Number 8 following equation (3). It was hence attempted to maintain y_n at $y_n = 1.5$ from $n = 21$ to 100 by solving the same optimization problem mentioned in Test 1, for each iteration of n. Results are plotted in Figure 3. Provided in the Figures below are only sample results. The experiment was repeated numerous times for consistency analysis with different patient models. Positive results were obtainable for all the cases. Full results are detailed in Appendix A [3] for interested readers.

$$\hat{y}_t = \sum_{i=1}^{u} a_i x_{n-i} + \sum_{j=1}^{v} b_j y_{n-j} + c \tag{2}$$

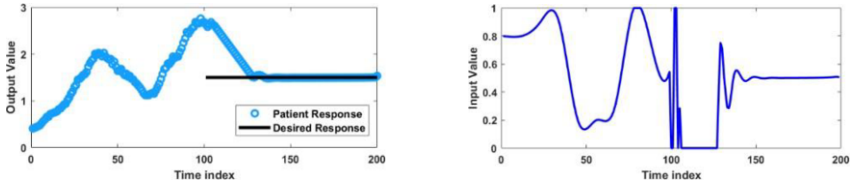
$$\hat{y}_t = \sum_{k=1}^{m} d_k DT^{(k)} + e \tag{3}$$

3. Results



(a) (b)

Figure 1. (a) Insulin trajectory (i.e., x_n) for patient Number 8 in the first 100 time instances; (b) Blood glucose response of patient Number 8 in the first 100 time instances alongside the estimated DT model for the large n case.



(a) (b)

Figure 2. (a) Controlling the blood glucose response of patient Number 8 in the large n case; (b) The trajectory of optimized insulin input for patient Number 8 in the large n case.

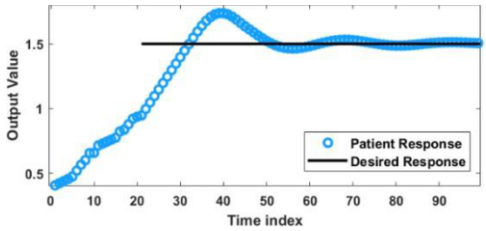


Figure 3. Controlling the blood glucose response of patient Number 8 in the small n case.

4. Discussion

Creating DTs of patients was attempted through synthetic data. Patients were modelled as time series. Therefore, our modelling is computationally efficient for real-time

decision support. Mathematical formulation of DTs was presented (in Appendix A) for two cases: (1) Having adequate amount of data (i.e., the large n case); and (2) Having inadequate amount of data (i.e., the small n case). The considered numerical example showed the ability to create DTs for both cases, even with a limited number of past patients (i.e., $k = 10$). As such, our objective of proof-of-concept is achieved. Our numerical analysis was limited only to two variables. One independent variable and one dependent variable. However, our theoretical formulation as presented in Appendix A is general to higher dimensional data. As such, there is potential for identifying and using DTs of patients in clinical practice.

5. Conclusions

Limitations of this work include some assumptions that we imposed. We assumed the availability of complete sets of numerical data that is equally spaced in time. However, complete numerical data equally spaced in time will be extremely rare in real- life clinical practice. Missing and incomplete data in healthcare are common [4]. Therefore, when going for real-world practice, significant data pre-processing will be required. This will involve interpolating missing or incomplete data (where possible) and converting any categorical data into numeric features [5]. Overcoming such issues will be critical to enable the use of DTs in healthcare. They will be challenging issues, but solvable.

In terms of modelling, there remains a lot of possibilities including time series modelling as we have attempted in this paper, to a whole range of other techniques spanning up to machine learning techniques.

Acknowledgments

This work is an outcome of the project titled “Technology and Dementia Care – A Conceptualization of Digital Twins for Facilitating Better Dementia Prediction, Prevention and Care” supported by the Barbara Dicker Brain Sciences Foundation of Swinburne University of Technology and an EMF grant from Epworth HealthCare awarded to the first author.

References

[1] Wickramasinghe, N., Jayaraman, P.P., Zelcer, J., Forkan, A.R.M., Ulapane, N., Kaul, R. and Vaughan, S., 2021. A vision for leveraging the concept of digital twins to support the provision of personalised cancer care. IEEE Internet Computing.

[2] Wickramasinghe, N., Ulapane, N., Andargoli, A., Ossai, C., Shuakat, N., Nguyen, T. and Zelcer, J., 2022. Digital twins to enable better precision and personalized dementia care. JAMIA open, 5(3), p.ooac072.

[3] Appendix A. Link: https://dit.swin.edu.au/DigitalTwin/20230309_MedInfo_Appendix_A_Formulation_of_DT.pdf

[4] Mirkes, E.M., Coats, T.J., Levesley, J. and Gorban, A.N., 2016. Handling missing data in large healthcare dataset: A case study of unknown trauma outcomes. Computers in biology and medicine, 75, pp.203-216.

[5] Golinko, E., Sonderman, T. and Zhu, X., 2017, June. CNFL: categorical to numerical feature learning for clustering and classification. In 2017 IEEE Second International Conference on Data Science in Cyberspace (DSC) (pp. 585-594). IEEE.