3K: Knowledge-Enriched Digital Twin Framework

Erkan Karabulut
e.karabulut@uva.nl
University of Amsterdam
The Netherlands

Paul Groth
p.t.groth@uva.nl
University of Amsterdam
The Netherlands

Victoria Degeler
v.o.degeler@uva.nl
University of Amsterdam
The Netherlands

Abstract

Digital Twins (DTs) are the digital equivalent of physical entities that facilitate, among others, monitoring and decision-making, thus helping extend the longevity of the twinned entity. DTs with automated decision-making capabilities require explainable inference mechanisms, especially for critical infrastructures such as water networks. Here we introduce 3K, a DT framework that aims for **k**nowledge-enriched inference that is explainable and fast, by synthesizing **k**nowledge representation (semantics) and **k**nowledge discovery methods. 3K constructs a knowledge graph, which is becoming a mainstream way of metadata storage in DTs, and proposes a new method that can run on both sensor data and knowledge graphs to learn semantic association rules. The rules represent the expected working conditions of the DT and we argue that when combined with domain knowledge in the form of ontological axioms, semantic association rules can help perform downstream tasks in DTs, including extending the longevity of the twinned entities such as an Internet of Things (IoT) system. Furthermore, we demonstrate the 3K framework in a water distribution network use case and show how it can be used for downstream tasks.

CCS Concepts

• **Information systems** → **Data mining**; **Association rules**; • **Computing methodologies** → **Rule learning**; **Neural networks**; • **Software and its engineering** → **Semantics**; • **Computer systems organization** → **Embedded and cyber-physical systems**.

Keywords

Digital Twin, Knowledge discovery, Neurosymbolic, Rule Learning, Semantic Web, Neural Networks

ACM Reference Format:
Erkan Karabulut, Paul Groth, and Victoria Degeler. 2024. 3K: Knowledge-Enriched Digital Twin Framework. In *14th International Conference on the Internet of Things (IoT 2024), November 19–22, 2024, Oulu, Finland.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3703790.3703834

1 Introduction

Digital Twin (DT) is a digital equivalent of a physical entities [7, 13] that has many successful applications across different domains including manufacturing [12], building management [30] and smart farming [24]. DTs aim to, among others, facilitate monitoring and decision-making and by doing so extend the longevity of the underlying system such as an Internet of Things (IoT) system [31]. The longevity, of an artifact, can be briefly defined as *fulfilling its intended purposes for a certain time, or as long as a defined set of conditions hold* [22]. In the context of cyber-physical systems where DTs have been frequently utilized [25], longevity relates to both software and hardware components.

In this paper, we propose the 3K framework, a knowledge-enriched decision-making framework for DTs that can be used in various tasks, including extending the longevity of the twinned system. Decision processes in DTs, especially in critical infrastructures where high-stake decisions are made, have high explainability requirements due to the associated risks, making black-box Artificial Intelligence (AI) models inadequate in these cases. Rule-based methods using Semantic Web technologies such as ontologies [8] and knowledge graphs [11] satisfy the explainability requirements [10]. However, they lack adaptation capabilities to learn from data as in Machine Learning (ML) methods and act accordingly. Therefore, 3K is a Neurosymbolic approach to explainable decision-making in DTs that utilizes both Semantic Web technologies and ML methods. It consists of 3 modules; i) **k**nowledge representation, ii) **k**nowledge discovery, and, iii) **k**nowledge-enriched inference, hence **3K**.

The knowledge representation part (Section 3.1) utilizes semantics, ontologies [8] and knowledge graphs [11], to represent static information in a DT, which is becoming a mainstream way of metadata storage in DTs [15, 19]. Semantic technologies add a layer of abstraction to the system and data model, thus decoupling them from further system updates which can help prolong its lifespan.

Knowledge discovery is the task of identifying meaningful patterns in the data [5] with many applications in the scope of DTs mainly to understand normal working conditions of certain components and processes and detect abnormalities based on their expected behavior [2, 4, 20]. Association Rule Mining (ARM), being a common knowledge discovery method that aims to learn associations between features of a given dataset in the form of logical rules [1], is not yet well-studied in DTs. The state-of-the-art ARM in DTs does not consider DT as a whole but only focuses on certain sub-components or tasks. As part of the knowledge discovery module (Section 3.2), we propose a new ARM method that runs on both static DT metadata and dynamic sensor data to learn semantic association rules that express patterns in DT data as a whole.

Lastly, the knowledge-enriched inference module (Section 3.3) performs reasoning over the learned semantic association rules and ontological axioms created by domain experts as part of the knowledge representation module. An example inference process in the scope of longevity is detecting system components running under unusual conditions to prevent future malfunctions and thus

IoT 2024, November 19–22, 2024, Oulu, Finland

ACM ISBN 979-8-4007-1285-2/24/11
https://doi.org/10.1145/3703790.3703834

extend their longevity. Working with rules satisfies the explainability requirements for high-stake decision making especially in DTs of critical infrastructures, e.g., to comply with certain requirements. It also allows easy integration of domain knowledge.

We demonstrate the 3K framework in a water distribution networks use case in Section 4 and emphasize its important aspects in Section 5. The goal of the 3K framework is not to propose a DT architecture, but a set of modules that DTs should have to support knowledge discovery over the entire DT data and make explainable inferences for various tasks including extending longevity.

2 Related Work

This section describes the related work on how semantics (knowledge representation), knowledge discovery, and knowledge-enriched inference are utilized in the scope of DTs.

2.1 Semantics in Digital Twins

Semantic technologies such as ontologies and knowledge graphs have been increasingly used in DTs [15, 19]. An ontology in computer science is defined as *'a formal, explicit specification of a shared conceptualization'* [8], while knowledge graphs are databases of structured semantic information often based on an ontology [11]. Semantics add a layer of abstraction over system and data models used in DT implementations. By doing so, they reduce the number of errors on the data interpretation level and thus can help extend the longevity of the systems. Other main objectives of semantics in DTs as reported in [15] are to establish semantic interoperability among various subcomponents and to infer semantic relations between DT data and components. However, semantics have not yet been widely used as part of learning and reasoning tasks in DTs, which is one of the goals of this paper.

2.2 Knowledge Discovery and Inference in Digital Twins

A common way of knowledge discovery is to apply data mining methods, specifically ARM, which refers to learning associations between features of a given dataset in the form of logical implications [1], e.g., $X \rightarrow Y$, which is read as *'if X then Y'*. X is called the antecedent of the rule while Y is called the consequent. In contrast to conventional applications of ARM, DTs include heterogeneous data from diverse sources including time series sensor data, and static metadata regarding DT system components. Association rules are typically mined from a single type of data, such as tabular [9] or graph data [23], and for a specific application [2] that are not able to provide a full view of how a DT works as a whole. Finding a suitable representation that can be used for ARM for a diverse set of data as in DTs is challenging. In addition, most of the state-of-the-art ARM algorithms struggle to run on big data [17, 26] which can be the case in large-scale DTs after combining various data sources.

The state-of-the-art knowledge discovery in DTs focuses on data integration, and learning patterns from specific system components or for a certain task. For instance, Mohammadi et al. [21] use knowledge discovery methods on smart city DTs that incorporate structured and unstructured data from multiple sources. A VR-based interactive user interface is provided to explore various aspects of the DT including health and environment. However, data mining methods are not utilized and reported as future works. Donkers et al. [4] combines semantic web technologies with knowledge discovery methods in a smart building DT use case. Semantic web technologies are used to collect and integrate data from various sources. A data analysis method described in [18] is used to analyze the impact of data features on the comfort of the building occupants. Liu et al. [20] utilized the Apriori [1] algorithm to mine association rules among risk factors in hoisting construction as part of a construction DT. Cai et al. [2] use FP-Growth [9] ARM algorithm in a DT of aircraft assembly process to detect quality deviations during the assembly.

Our approach. In contrast to the existing knowledge discovery methods applied as part of DTs, we propose to run ARM algorithms on both sensor data and knowledge graphs describing the context in which the sensor data is obtained, hence both static and dynamic parts of DT data. In this way, we aim to obtain a full overview of how the DT system should behave according to the mined patterns. In addition, we propose a DL-based ARM approach that can learn a concise set of high-quality semantic association rules. To the best of our knowledge, there has been no study that combines ontological axioms and semantic association rules for inference or decision-making in DTs.

3 3K: Knowledge-Enriched Digital Twins

This section describes the 3K framework that aims for explainable inference in DTs. Figure 1 illustrates the entire pipeline of operations for the proposed 3K framework.

3.1 Knowledge Representation

Knowledge representation in DTs refers to the semantic modeling of the DT system and its components, often in the form of a knowledge graph with an underlying domain ontology. The domain ontology contains the vocabulary used in the domain to which the Physical Twin (PT) belongs. In addition, the ontology may also contain a set of logical axioms that represent constraints over the system components as well as sensor data. One simple example from the water distribution network domain is *when two pipes are attached to each other and have no other connections except on the opposite sides of the pipes, the water flow in both pipes has to be the same.* These constraints representing a specific domain knowledge, together with association rules learned from DT data, will be used in the knowledge-enriched inference module for discrepancy detection.

In this module, a knowledge graph is constructed from DT system metadata (asset descriptions) and a domain ontology via semantic matching. The domain ontology contains the vocabulary used in the domain to which the physical twin belongs. Asset descriptions refer to the files that contain metadata about the system components. The goal of semantic matching is to map asset descriptions to the classes, relations, and properties in the domain ontology. Then, the obtained mappings correspond to a knowledge graph describing all the assets in the system.

One example of semantic matching in the literature is done via Natural Language Processing (NLP) methods [16] based on the semantic descriptions (can be regarded as asset descriptions) of IoT devices. Another way of matching, when the number of concepts

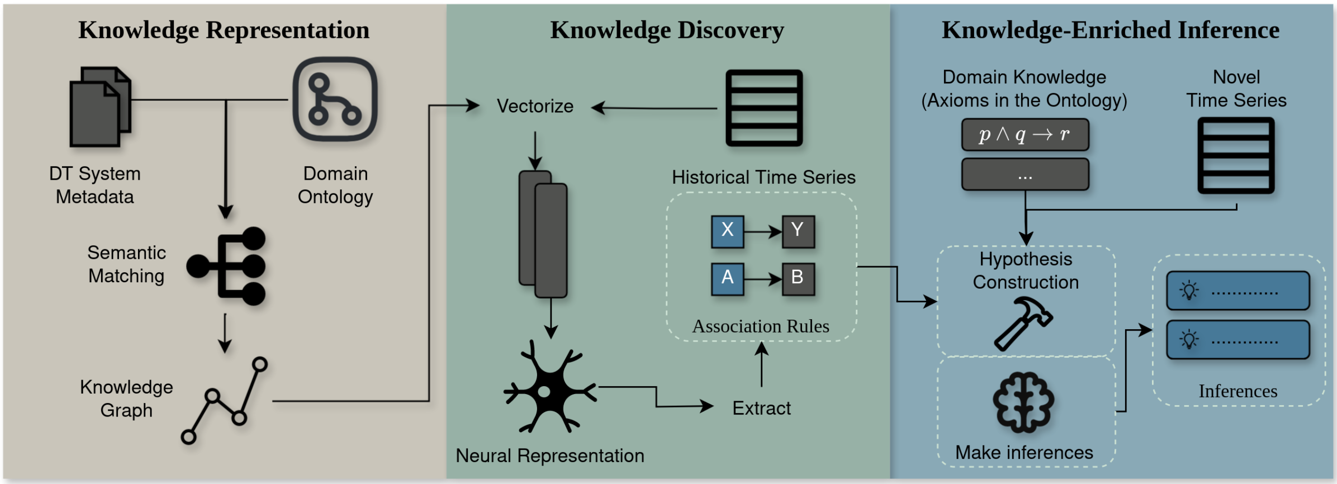

Figure 1: 3K: knowledge representation, knowledge discovery, and knowledge-enriched inference modules.

in the ontology is low, is by pre-defining a set of mappings between the asset descriptions and the ontological elements.

3.2 Knowledge Discovery

We propose an ARM algorithm that utilizes dynamic sensor data and static system metadata represented as a knowledge graph for knowledge discovery in DTs. The role of knowledge graphs is to provide context to the sensor data during the rule learning process. In this way, we aim to discover patterns not only based on sensor measurements but also on the context in which they are placed.

An association rule learned from sensor data looks as follows: *'if sensor1 measures a value in range R1, then sensor2 must measure a value in range R2'*. On the other hand, association rules with semantics from knowledge graphs are more generically applicable and more explainable: *'if a sensor S placed inside component C that has a set of semantic properties S measures a value in range R1, then sensor S2 placed inside component C2 with semantic properties S2 will measure a value in range R2.'* This second rule relates to not only *sensor1* and *sensor2* as in the first rule, but it rather describes a certain context and measurements taken in that context which makes it more generically applicable. The semantic properties may correspond to the components C and C2, or their neighbors on the knowledge graphs.

As one way to implement the proposed method, we developed a DL-based ARM approach which can be found in [14]. The proposed method is illustrated in Figure 3. Combining sensor data with knowledge graphs may result in big high-dimensional data in the case of large-scale DTs. As the state-of-the-art ARM methods struggle on big-high dimensional data [17, 26], a DL-based ARM method that makes use of denoising Autoencoders [27] is proposed. An Autoencoder is a neural network architecture that creates a lower dimensional representation of a given input referred to as the code layer, and then reconstructs the input from the code layer.

First, sensor data is enriched by the corresponding semantic properties from the knowledge graph. Each sensor has a representation on the knowledge graph and the binding refers to the mapping of sensor data to the corresponding node on the knowledge graph. Semantic enrichment refers to the coupling of sensor data with the properties of the sensor node itself, the properties of the node that describes where the sensor is placed, and/or the neighbors of those nodes in the knowledge graph which would provide a wider context. Second, the semantically enriched sensor data is vectorized by discretization and one-hot encoding. The third step is to create a neural representation of the vectors using an under-complete denoising Autoencoder. In the output layer, the Autoencoder has probability distributions per input feature. We leverage the reconstruction feature of Autoencoders to learn and extract associations between input features. As the last step, semantic association rules are extracted from the neural representation as described below.

Rule extraction example. Figure 2 illustrates a rule extraction example from a trained Autoencoder. Assume that a denoising Autoencoder is trained on 2 features: $f_1 = \{a, b\}$, $f_2 = \{c, d, e\}$. To test whether f_1 being a implies f_2 being c, a *test vector* with equal probabilities per class value is created ($[0.5, 0.5, 0.33, 0.33, 0.33]$), and then $f_1(a)$ is marked with 100% probability: $[1, 0, 0.33, 0.33, 0.33]$ (the first 2 numbers correspond to f_1 values and the remaining 3 correspond to f_2 values). Assume that a forward run on the trained Autoencoder returns $[0.97, 0.03, 0.01, 0.05, 0.94]$. Since the probability of $f_2(e)$ is high (higher than a given threshold, e.g., 90%),

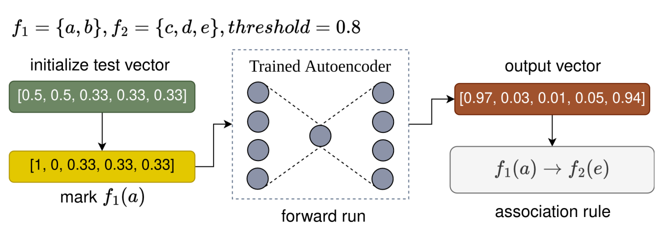

Figure 2: An association rule extraction example from a trained Autoencoder.

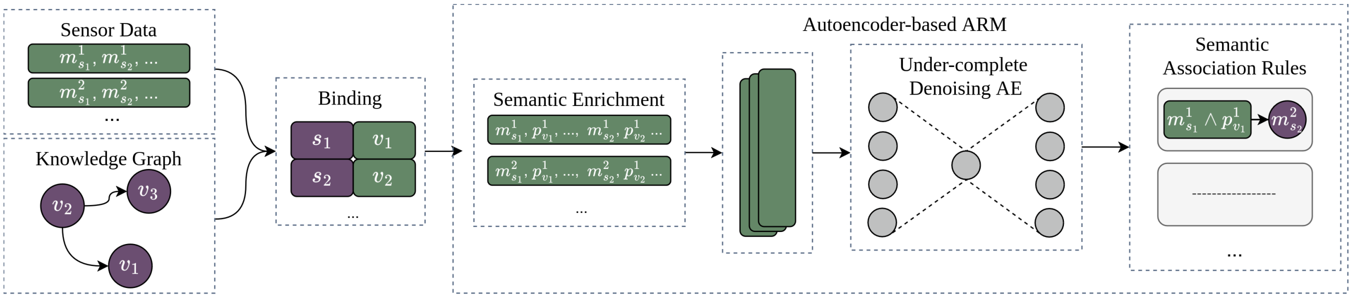

Figure 3: Learning semantic association rules from sensor data and knowledge graphs [14].

we conclude that $f_1(a) \longrightarrow f_2(e)$. The full algorithm to create test vectors can be found in [14].

3.3 Knowledge-Enriched Inference

This section describes making inferences with the learned semantic association rules and domain knowledge represented as axioms of a domain ontology. In comparison to black box DL-based inference, we hypothesize that this form of inference is explainable and faster, as checking whether a rule holds is a simple comparison operation.

Preprocessing rules. Depending on the inference task, a rule quality metric that applies to the task is selected and low-quality rules are filtered out. A plethora of rule quality metrics are proposed in ARM literature [17]. Next, the consistency of the selected rules is checked with a logical reasoner based on the axioms defined in the ontology, and inconsistent rules are eliminated.

Constructing hypotheses. The pre-processed rules are then grouped together to form so-called *hypotheses* which refers to normal working conditions of the DT. We identified 2 potential directions to form hypotheses. In the case of having labeled abnormal data, the first approach relates to learning associations between the rules and abnormalities. We aim to associate a certain subset of rules with a certain abnormality, or in general to an abnormality, hence forming hypotheses. We call these abnormalities *discrepancies* between PT and DT since they refer to a part of physical space that is not yet reflected in the DT. The second is a neuro-symbolic approach in which the rules are grouped based on their similarity in the case of not having labeled abnormal data. In this case, grouping can be based on the applicability of the rules on certain components in the DT or the overlap among themselves.

Inference time. Hypotheses are checked on novel unseen data, and we argue that the hypotheses that do not hold for a certain time act as a good indication that a discrepancy exists at a certain part of the DT system. Another example application of the hypotheses is to understand whether a system update, e.g., a software or hardware update, will cause disruptions by checking their validity after the update in a controlled environment. The rules and the hypotheses are periodically updated based on the incoming sensor data and system changes such as static data changes on the knowledge graph.

4 Industrial Example: Drinking Water Distribution Networks

The proposed framework has been implemented as a part of a DT of a water distribution network [3], with the exception of the knowledge-enriched inference module which is currently under development, and the source code can be found in: **https://github.com/DiTEC-project/semantic-association-rule-learning**.

4.1 Knowledge graph construction

Conventional DTs in the water distribution networks domain utilize physical simulation tools such as EPANET [1] to model the system and test various scenarios. LeakDB [29] and L-Town [28] are two artificially generated realistic water distribution network datasets that are based on EPANET. The datasets contain *.inp* files which describe the simulated water network and provide metadata (can be seen as asset descriptions).

At the time of writing this paper, no ontologies are available for the domain of water distribution networks. Therefore, we utilized the data schema of the EPANET software to map asset descriptions to the vocabulary described as part of the data schema in a rule-based fashion as the number of concepts is low. Figure 4 shows an excerpt from the knowledge graph constructed using the LeakDB dataset. The figure shows nodes of different types including pipes and junctions, where each node contains a set of properties such as the diameter of a pipe that is shown on the right side of the figure.

[1]https://www.epa.gov/water-research/epanet
[2]https://neo4j.com/

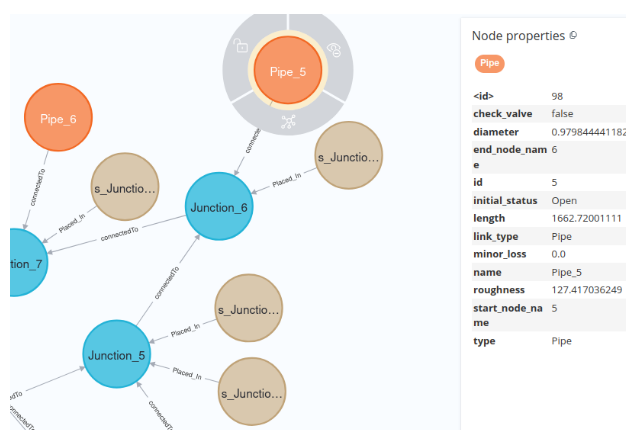

Figure 4: An excerpt of a knowledge graph constructed from the LeakDB dataset on Neo4j user interface[2].

Table 1: Association rule examples with (top) and without (bottom) semantics learned from LeakDB dataset.

Association Rule	Support	Coverage
if a water flow sensor s1 is inside a Pipe p1 with length 843-895, and a water demand sensor s2 inside a Junction j1 that is connected to p1 measures 13-17, then s1 must measure between 23-31.	0.5	0.54
if the water flow sensor inside Pipe_28 measures between 23-31, then the water flow sensor inside Pipe_18 must measure between -767−471.	0.43	0.52

4.2 Semantic association rule learning

A set of semantic association rules is learned based on the knowledge graph constructed in the previous step and the sensor data using the proposed method in Section 3.2. The evaluation is based on the rule quality criteria that are commonly used in the ARM literature such as support, confidence, coverage, Zhang's metric, and execution time [17]. The quality of the rules is compared extensively with the state-of-the-art ARM methods and discussing the results is beyond the scope of this paper, also due to the space restrictions. Please refer to [14] for the extensive evaluation.

Table 1 shows two association rules examples learned from the LeakDB dataset. The bottom row shows an association rule learned from sensor data only, which describes a certain condition that is applicable to two sensors specifically, namely the water flow sensors inside Pipe_28 and Pipe_18. On the other hand, the top row shows a semantic association rule that is learned from both knowledge graphs and sensor data. It describes a certain pattern that is independent of individual sensors and is applicable when the antecedent of the rule holds. We say that the first rule is more explainable as it provides more context via the semantic properties as opposed to the second rule which only includes sensor measurements.

The support value is the percentage of transactions (data instances) with a certain item or rule among all transactions, while the coverage refers to the percentage of transactions for which the rules are applicable. As both support and coverage values are higher for the semantic association rule given in the first row, we conclude that the semantic association rules are more generically applicable. We argue that this will result in requiring less number of rules in the knowledge-enriched inference step, to have full coverage over the dataset, and thus speeding up the inference further.

4.3 Fast and Explainable Inference in Digital Twins

This module is currently under development. As described in Section 3.3, the semantic association rules are used with ontological axioms to make inferences. As an example, if the rules listed in Table 1 do not hold for a certain period of time, this may indicate a water leakage in one of the pipes. Since there are no open-source water distribution network ontologies, we are collaborating with domain experts at a water distribution company, Vitens[3], to create an ontology. We argue that the semantic association rules together with the ontological axioms represent the expected working conditions of a water distribution network.

As a first step, we plan to detect abnormalities in the system such as water leakages and sensor degradations. Besides the statistical evaluation of the semantic association rules with rule quality criteria, their capability to detect abnormalities will also be used in further evaluation of our proposed rule learning approach. The proposed approach will be compared with both other rule-based approaches as well as other methods that are both task-specific such as leak detection methods and other more generic abnormality detection methods. Furthermore, the evaluations will be repeated on datasets from different domains such as the energy domain, as also performed in [14] for semantic association rule learning on LBNL fault detection dataset [6].

5 Discussion

This section discusses important aspects of the 3K framework.

Knowledge-enriched Digital Twins. Semantics technologies have been increasingly used in DTs for various tasks as pointed out in recent literature reviews [16, 19]. Despite providing highly valuable information regarding the DT systems, semantics are not utilized as part of learning tasks in DTs. Our proposed 3K framework includes the first case where semantics, i.e. knowledge graphs, makes a significant improvement in a learning task in DTs, namely learning association rules, from sensor data by learning rules that are more generically applicable and reflect DT systems as a whole rather than sensor data only. We argue that there is a huge potential in the direction of utilizing semantics in DTs to facilitate learning.

Explainable decision-making with Neurosymbolic AI. Knowledge discovery methods such as ARM are used in DTs to detect abnormalities in DTs as they are explainable methods to decision-making. However, the existing applications of ARM in DTs give limited consideration to the characteristics of DT data such as heterogeneity and high dimensionality. DTs of large-scale systems can have a high number of sensors with various types deployed which are treated as different data dimensions in ARM algorithms. Moreover, state-of-the-art ARM methods are developed to run on a single type of data such as tabular data, time series data, or graph data. We argue that neural networks can help creating a common neural representation of diverse datasets as in DTs in an efficient way as they are also capable of handling big data. Therefore, in the knowledge discovery module of our 3K framework, we provided the first method in DTs to create a neural representation of a DT knowledge graph and sensor data using Autoencoders, and extracted association rules from the neural representation algorithmically, making it a neurosymbolic approach.

Alternative neural representations of Digital Twin data. 3K framework utilizes an Autoencoder to create a neural representation of the DT data. However, this is only an initial step and other neural network architectures are to be experimented with for their capability of representation learning from DT data. As an example, Graph Neural Networks (GNNs) are better at capturing graph-structured data, and new methods to integrate GNNs

[3] https://www.vitens.nl/

and sensor data can be explored for better representation learning. Other architectures are likely to necessitate new ways of extracting associations from neural representations, which we also see as a future research direction.

6 Conclusion and Future Work

This study proposed the 3K framework that aims for explainable decision-making in DTs which can be used for various tasks including abnormality detection that potentially extends the longevity of the twinned system. It consists of knowledge representation, knowledge discovery, and knowledge-enriched inference modules. The knowledge representation module constructs a knowledge graph from DT asset descriptions and a domain ontology by semantic matching. Semantic association rules are learned from sensor data and the knowledge graph using a Deep Learning (DL)-based approach to have a full overview of the behavior of the DT in the knowledge discovery module. The proposed approach utilizes Autoencoders to create a neural representation of the DT data, and extracts semantic association rules from the neural representation.

In future work, as part of the knowledge-enriched inference, we plan to perform reasoning over the semantic association rules, and ontological axioms to detect discrepancies between the DT and its physical counterpart. Furthermore, we plan to evaluate our proposed approach on various tasks across different domains to further validate its efficiency alongside the statistical rule quality evaluation. This includes both application-level tasks such as leakage detection in water networks or fault detection in the energy domain, as well as systems' longevity-related tasks such as understanding how would a new hardware or software update affect the system based on the previously learned patterns and domain knowledge.

Acknowledgement: This work is funded by the project DiTEC: Digital Twin for Evolutionary Changes in Water Networks (NWO 19454).

References

[1] Rakesh Agrawal, Ramakrishnan Srikant, et al. 1994. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, Vol. 1215. Santiago, 487–499.

[2] Hongxia Cai, Jiamin Zhu, and Wei Zhang. 2021. Quality deviation control for aircraft using digital twin. *Journal of Computing and Information Science in Engineering* 21, 3 (2021), 031008.

[3] VO Degeler, Mostafa Hadadian, Erkan Karabulut, and Alexander Lazovik. 2024. DiTEC: Digital Twin for Evolutionary Changes in Water Distribution Networks. In *International Symposium On Leveraging Applications of Formal Methods, Verification and Validation (ISoLA)*.

[4] Alex Donkers, Bauke de Vries, Dujuan Yang, P Pauwels, M Poveda-Villalón, and W Terkaj. 2022. Knowledge Discovery Approach to Understand Occupant Experience in Cross-Domain Semantic Digital Twins.. In *LDAC@ ESWC*. 77–86.

[5] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. 1996. From data mining to knowledge discovery in databases. *AI magazine* 17, 3 (1996), 37–37.

[6] Jessica Granderson, Guanjing Lin, Yimin Chen, Armando Casillas, Piljae Im, Sungkyun Jung, Kyle Benne, Jiazhen Ling, Ravi Gorthala, Jin Wen, Zhelun Chen, Sen Huang, , and Draguna. Vrabie. 2022. LBNL Fault Detection and Diagnostics Datasets. https://doi.org/10.25984/1881324

[7] Michael Grieves. 2014. Digital twin: manufacturing excellence through virtual factory replication. *White paper* 1, 2014 (2014), 1–7.

[8] Tom Gruber. 1993. What is an Ontology.

[9] Jiawei Han, Jian Pei, and Yiwen Yin. 2000. Mining frequent patterns without candidate generation. *ACM sigmod record* 29, 2 (2000), 1–12.

[10] Timon Hoebert, Wilfried Lepuschitz, Erhard List, and Munir Merdan. 2019. Cloud-based digital twin for industrial robotics. In *Industrial Applications of Holonic and Multi-Agent Systems: 9th International Conference, HoloMAS 2019, Linz, Austria, August 26–29, 2019, Proceedings 9*. Springer, 105–116.

[11] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutiérrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan F. Sequeda, Steffen Staab, and Antoine Zimmermann. 2021. *Knowledge Graphs*. Number 22 in Synthesis Lectures on Data, Semantics, and Knowledge. Springer. https://doi.org/10.2200/S01125ED1V01Y202109DSK022

[12] Wenjie Jia, Wei Wang, and Zhenzu Zhang. 2023. From simple digital twin to complex digital twin part II: Multi-scenario applications of digital twin shop floor. *Advanced Engineering Informatics* 56 (2023), 101915.

[13] David Jones, Chris Snider, Aydin Nassehi, Jason Yon, and Ben Hicks. 2020. Characterising the Digital Twin: A systematic literature review. *CIRP journal of manufacturing science and technology* 29 (2020), 36–52.

[14] Erkan Karabulut, Victoria Degeler, and Paul Groth. 2024. AE SemRL: Learning Semantic Association Rules with Autoencoders. *arXiv preprint arXiv:2403.18133* (2024). https://arxiv.org/abs/2403.18133

[15] Erkan Karabulut, Salvatore F Pileggi, Paul Groth, and Victoria Degeler. 2023. Ontologies in digital twins: A systematic literature review. *Future Generation Computer Systems* (2023).

[16] Erkan Karabulut and Rute C Sofia. 2023. An Analysis of Machine Learning-Based Semantic Matchmaking. *IEEE Access* 11 (2023), 27829–27842.

[17] Minakshi Kaushik, Rahul Sharma, Iztok Fister Jr, and Dirk Draheim. 2023. Numerical association rule mining: a systematic literature review. *arXiv preprint arXiv:2307.00662* (2023).

[18] Jeehee Lee and Youngjib Ham. 2021. Physiological sensing-driven personal thermal comfort modelling in consideration of human activity variations. *Building Research & Information* 49, 5 (2021), 512–524.

[19] Franz Georg Listl, Daniel Dittler, Gary Hildebrandt, Valentin Stegmaier, Nasser Jazdi, and Michael Weyrich. 2024. Knowledge Graphs in the Digital Twin: A Systematic Literature Review About the Combination of Semantic Technologies and Simulation in Industrial Automation. *arXiv preprint arXiv:2406.09042* (2024).

[20] Zhansheng Liu, Xintong Meng, Zezhong Xing, and Antong Jiang. 2021. Digital twin-based safety risk coupling of prefabricated building hoisting. *Sensors* 21, 11 (2021), 3583.

[21] Neda Mohammadi and John Taylor. 2020. Knowledge discovery in smart city digital twins. (2020).

[22] Diogo Proenca, Goncalo Antunes, Jose Borbinha, Artur Caetano, Stefan Biffl, Dietmar Winkler, and Christoph Becker. 2013. Longevity as an Information Systems Design Concern.. In *CAiSE Forum*. 73–80.

[23] Lucia Sacchi, Cristiana Larizza, Carlo Combi, and Riccardo Bellazzi. 2007. Data mining with temporal abstractions: learning rules from time series. *Data Mining and Knowledge Discovery* 15 (2007), 217–247.

[24] Petr Skobelev, Aleksey Tabachinskiy, Elena Simonova, and Oleg Goryanin. 2022. Development of crop-simulation multiagent system for smart digital twin of plant. In *2022 VIII International Conference on Information Technology and Nanotechnology (ITNT)*. IEEE, 1–8.

[25] Fei Tao, Qinglin Qi, Lihui Wang, and AYC Nee. 2019. Digital twins and cyber–physical systems toward smart manufacturing and industry 4.0: Correlation and comparison. *Engineering* 5, 4 (2019), 653–661.

[26] Akbar Telikani, Amir H Gandomi, and Asadollah Shahbahrami. 2020. A survey of evolutionary computation for association rule mining. *Information Sciences* 524 (2020), 318–352.

[27] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*. 1096–1103.

[28] SG Vrachimis, DG Eliades, R Taormina, A Ostfeld, Z Kapelan, S Liu, MS Kyriakou, P Pavlou, M Qiu, and M Polycarpou. 2020. Dataset of BattLeDIM: Battle of the leakage detection and isolation methods. In *Proc., 2nd Int CCWI/WDSA Joint Conf. Kingston, ON, Canada: Queen's Univ.*

[29] Stelios G Vrachimis, Marios S Kyriakou, et al. 2018. LeakDB: a Benchmark Dataset for Leakage Diagnosis in Water Distribution Networks:(146). In *WDSA/CCWI Joint Conference Proceedings*, Vol. 1.

[30] Xiang Xie, Jorge Merino, Nicola Moretti, Pieter Pauwels, Janet Yoon Chang, and Ajith Parlikad. 2023. Digital twin enabled fault detection and diagnosis process for building HVAC systems. *Automation in Construction* 146 (2023), 104695.

[31] Peter Zdankin, Marco Picone, Marco Mamei, and Torben Weis. 2022. A digital-twin based architecture for software longevity in smart homes. In *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 669–679.