Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

EXPLAINABLE MODELING IN DIGITAL TWIN

Lu Wang
Tianhu Deng

Department of Industrial Engineering
Tsinghua University
Beijing, 100084, CHINA

Zeyu Zheng
Zuo-Jun Max Shen

Department of Industrial Engineering
and Operations Research
University of California, Berkeley
Berkeley, CA 94720, USA

ABSTRACT

Stakeholders' participation in the modeling process is important to successful Digital Twin (DT) implementation. The key question in the modeling process is to decide which options to include. Explaining the key question clearly ensures the organizations and end-users know what the digital models in DT are capable of. To support successful DT implementation, we propose a framework of explainable modeling to enable the collaboration and interaction between modelers and stakeholders. We formulate the modeling process mathematically and develop three types of automatically generated explanations to support understanding and build trust. We introduce three explainability scores to measure the value of explainable modeling. We illustrate how the proposed explainable modeling works by a case study on developing and implementing a DT factory. The explainable modeling increases communication efficiency and builds trust by clearly expressing the model competencies, answering key questions in modeling automatically, and enabling consistent understanding of the model.

1 INTRODUCTION

Digital Twin (DT) is a set of digital models and supporting data to represent the real world and offer intelligent decisions in complex systems (Digital Twin Consortium 2020; Shao et al. 2019; Stark and Damerau 2019; Wang et al. 2020). Building a DT system involves model developers (hereafter modelers) and a range of potential stakeholders who are interested in the model or are affected by the insights from the model. The most common roles include (1) commissioners who specify the general purpose and scope of the model, provide required resources for the modeling, and monitor the use of the model; (2) users who own the problem that the model is designed to solve and usually are domain experts (Calder et al. 2018).

Building trust in DT is a critical challenge for its successful implementation. Digital models used in DT should be explained clearly to ensure the commissioners and users know what DT is capable of (Fuller et al. 2020). The low stakeholder participation in the modeling process results in a significant limitation in DT implementation (Jahangirian et al. 2015; Calder et al. 2018; Taillandier et al. 2019). In particular, the DT concept has been widely used in smart buildings and cities (Du et al. 2020; Pan and Zhang 2021; Lu et al. 2020) and is viewed as "a new game-changing approach to construction automation ... that will transform the industry quicker than ever before", according to UNSW Newsroom (2020). Construction project management especially emphasizes the importance of good team communication, since poor communication endangers trust and might lead to a considerable loss (Du et al. 2020; Cerić 2015).

We consulted several papers on how modelers are instructed to work with other stakeholders to determine the key question in modeling: which option should be included and which not in the digital model. Robinson (2015) suggested performing the following activities, including understanding the problem

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

situation, determining the general project objectives, identifying inputs and outputs of the digital model, determining the scope and fidelity of the model, identifying any assumptions and simplifications. Calder et al. (2018) summarized a checklist on making and using models. The checklist contains questions to clarify the understanding of what will be involved, such as purpose, scope, output and follow-up, and resources. Most of the current instructions are a series of open questions. It will be ineffective if the communication fails to send a clear message about the project goals, objectives, and progress (Levasseur 2010).

Conventionally, the interaction between modelers and stakeholders in the modeling process is a top-down, one-way communication (Levasseur 2010). The specification and expectations are handed over to modelers. Then, modelers design, build and validate the model and return the results to the stakeholders (Calder et al. 2018). The one-way communication can lead to low stakeholder participation, which is among the top-ranked nontechnical causal factors in project failure (Levasseur 2010).

To address the above problems, this paper proposes a conceptual framework of explainable modeling since explanations prove to improve trust (Shin 2021). As far as we know, our research is the first to propose explainable modeling to support stakeholders' participation in the modeling and simulation research. This paper intends to answer the following questions: how to clearly and explicitly express the modeling process? what types of explanations are needed to improve trust? and how to measure the value of explainable modeling?

To answer the above questions, we begin in Section 2 with a motivating case. Section 3 introduces our explainable modeling framework. We first formulate the modeling process mathematically to clearly express the modeling process. Then, to answer the key question of which option should be included and which not in the digital model, we develop three types of explanation, including model-based explanation, scenario-based explanation, and goal-oriented explanation. In Section 4, three explainability scores are proposed to measure the value of the explainable modeling framework. In Section 5, we present a case study based on the development and implementation of a DT factory to illustration our framework. We conclude in Section 6.

2 A MOTIVATING CASE

We begin the discussion with a motivating case. We collaborated with a leading consumer packaged goods (CPG) manufacturing company to develop a DT factory. The name and specific market segment of the manufacturer cannot be disclosed due to confidentially concerns. The manufacturer has over 100 SKUs (e.g., per pack and per carton with different brands) and serves more than 300 regional markets in China by collaborating with 33 distributors. The annual turnover reaches RMB 90 billion in 2019. In this case, we play the role of modelers; the commissioner is the factory manager; and the users include the information system engineers, the domain experts, and workers in the factory.

The ultimate goal of this project is to build a DT factory to achieve advanced digitalization and maintain a competitive advantage in the volatile market. One of the important tasks in the DT project is the data integration problem (Jones et al. 2020). To mirror the state of the factory, data should be collected from the physical factory or integrated from various information systems. There are multiple data sources (options for the digital model). The modeling is to determine which data (option) should be included and which not. The integration problem leads to a series of checks during the modeling process. First, how the decision quality will be improved by including the option. Second, how the model and computation will be complicated by including the option. Third, how the integration cost will be increased by including the option. For ease of illustration, we specify the decision quality as the solution quality of the production planning and scheduling in our case study.

Understanding the problem situation cannot be just left to the modelers but requires multidisciplinary expertise (Calder et al. 2018). We, as the *modelers*, know how the option influences the model (size and computational complexity) and solution quality. The information systems engineers are experts in evaluating the cost of collecting data from the factory and various systems. The domain experts are familiar

with the business scenarios and traditional manual scheduling rules. They are also helpful in evaluating the solution quality.

Explainable modeling is required to communicate ideas and knowledge clearly (Lupeikiene et al. 2014). We summarize the complexity of the practical problem in Table 1. Imagine that conventional one-way communication is applied. Modelers develop comprehensive models of resource requirements for automatically-generated production schedules to realize smart manufacturing, while users, who are absent from the modeling process, will find the underlying scheduling logic are too complex and difficult to understand and thus proceed with their traditional manual scheduling approaches (Techtarget 2008). Therefore, explainable modeling is required to highlight the key information and increase stakeholders' understanding of the modeling process.

Table 1: Complexity of the practical problem.

Data	Description
Number of items	31
Number of machines and machine groups	47,21
Number of shift per day	3
Number of sequence-dependent changeover time	372

3 EXPLAINABLE MODELING

In this section, we answer the questions of how to clearly and explicitly express the modeling process and what types of explanations are needed to improve trust. In Section 3.1, we formulate the modeling process mathematically. Next, we explore the desirable properties of a good explanation and propose three approaches to generate explanations automatically in Section 3.2.

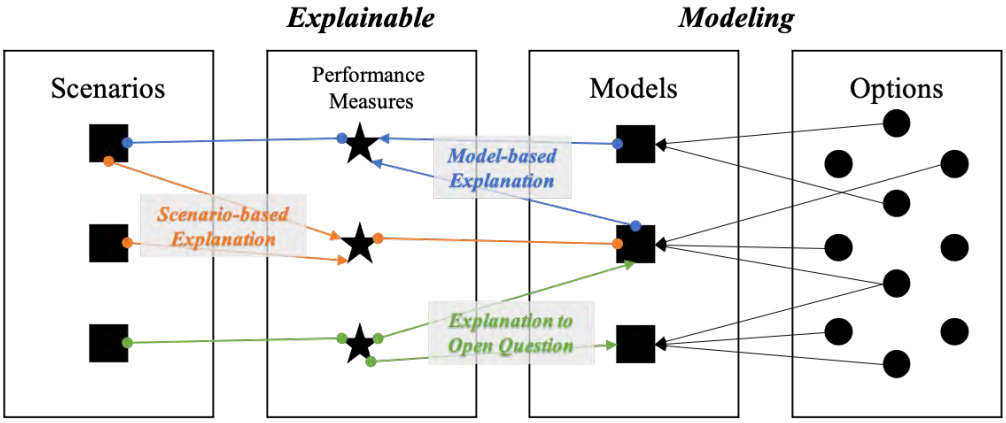

Figure 1: The framework of explainable modeling.

The framework of explainable modeling is summarized in Figure 1. The overall goal of the modeling process is to choose the best model (in other words, to determine which options to include) that satisfies the performance measures under certain scenarios. In practice, the performance measures are often a set of competing objectives that are influenced by both the scenarios and models. Therefore, the proposed three types of explanation generation approach launch around performance tradeoffs under different scenarios and models. For example, in the above case, the choice of data sources will lead to digital models of the factory with different levels of fidelity. Both the model complexity and solution quality are considered important performance measures in DT implementation. The business scenarios, including product variety and demand volume, have a significant influence on the solution quality as well. The explanations, in this

case, explain how the data sources inclusion and business scenarios influence the model complexity and solution quality.

3.1 Formulating the Modeling Process

To clearly and explicitly express the modeling process, we formerly formulate the underlying concepts, including options, models, scenarios, and performance measures.

Options and models. We consider a configurable system with a set of all candidate options, denoted by $\mathscr{O}$. We formulate a digital model $m \in \mathscr{M}$ in DT as a function such that $m : \mathscr{O} \to \{0,1\}$. If the option o is included in the model m, then we have $m(o) = 1$; otherwise, we have $m(o) = 0$. In order to compare different models, we define $\Delta(m_1, m_2) : \mathscr{M} \times \mathscr{M} \to \mathscr{O}$ as the function that returns the additional options included in m_1, and we have $\Delta(m_1, m_2) = \{o \in \mathscr{O} : m_1(o) = 1, m_2(o) = 0\}$. We do not specific the options in the formulation. The specification of options is determined by application domain and should be understandable to all parties in the project.

The formulation approach of the options and models is inspired by Siegmund et al. (2015). They develop a method to describe the influences of configuration options and their interactions. They use linear regression to learn and estimate the individual options and their interactions' influence on model performance. In particular, the proposed method is used to balance the tradeoffs between model accuracy, model size, and computation efforts (Kolesnikov et al. 2019). Our work differs from theirs in two aspects. First, we do not restrict a linear relationship between the options and the model performance. Second, we consider the impact of business scenarios on model performance which is introduced below.

Scenarios and performance measures. In addition to the selection of options, model performance is also context-sensitive (Yilmaz and Liu 2020) and have to support explanation or exploration of future scenarios (Calder et al. 2018). Let $\mathscr{S}$ be the set of all business scenarios. A scenario s can include many instances. An instance represents a realization of data inputs required in the model m.

The construction of a model must balance competing objectives (Bertsimas and King 2016; Kolesnikov et al. 2019). We consider a set of performance measures, labeled by i. A performance measure is formulated as a function such that $\pi_i : \mathscr{M} \times \mathscr{S} \to \mathbb{R}$. The performance measure can be used to evaluate (1) solution quality (e.g., the accuracy of a predictive model, the demand satisfaction rate of an inventory planning model); (2) model configuration (e.g., model size, running time). For simplicity, in this section, we assume that the higher the value is, the better the performance measure. We relax this assumption in the case study.

An illustrative example. We take the motivating case in Section 2 as an example to illustrate our formulation. We have $m(o_{change}) = 1$ if the option of machine changeover time, denoted as o_{chnge}, is included in the digital model for the factory. Let s denote a production scenario where the product variety is high and the production volume of each product is low. The model is supposed to balance several competing objectives, such as running time and the machine utilization rate. Each objective can be formulated by a performance measure in our formulation. For instance, if the running time of the model m given the business scenario s is 90 seconds, we have $\pi_{run}(m, s) = 90$.

Rationale of our formulation. We conclude three reasons that motivate us to formulate the modeling process mathematically. First, a mathematical formulation can clearly and explicitly express what options are included in the model and how the option selection and scenarios influence the performance measures. Without a quantitative model, people often create an imperfect mental model to understand the complex system due to the cognitive limitations (Norman 1988; Norman 1983). A clear expression avoids misunderstanding caused by biased mental models. Having a consistent understanding of what is involved and what is expected improves communication efficiency and builds interpersonal trust (Harper, Mustafee, and Yearworth 2021).

Second, a mathematical formulation enables composability that is "the capability to select and assemble simulation components in various combinations into valid simulation systems to satisfy specific user requirements" (Weisel et al. 2003). As the model complexity grows, the model could be misused or misunderstood and, much more often, it remains uncertain about what options should be included in a

model (Calder et al. 2018). Our formulation considers the impact of options and business scenarios separately and helps modelers and other stakeholders work closely to uncover the implicit causality in the problem. The right option can be selected to improve the particular performance measures under a specific business scenario, which increases model composability.

Third, our formulation has compatibility with analytics, such as network analysis and optimization problems. For example, the options and scenarios can be formulated as a network and analyzed by the automatic structural analysis proposed by Honti et al. (2019). They use the network representation to highlight the relationship between different items and measure the importance of each item. Recall that we use building a multivariate regression model as an example. Jointly addressing competing objectives in such modeling tasks can be solved by a mixed-integer quadratic optimization proposed by Bertsimas and King (2016). However, applying their approach directly in our formulation may lead to intractable problems, since our formulation can deal with more general modeling tasks where convexity does not necessarily hold.

3.2 Generating Explanation

In the modeling process, it is critical to understand how different options and scenarios influence the model performances and which options should be included. Therefore, we propose the following three types of explanation that closely launch around the above key questions. First, model-based explanation focuses on explaining how the option selection influences the performance measures. Second, scenario-based explanation focuses on explaining how the model performances change under various business scenarios (data input). Third, goal-oriented explanation focuses on explaining how to modify the model to improve a certain performance measure.

It is important to notice that a good explanation should be contrastive (Miller 2019). For example, the commissioner and user would not ask why model m is proposed but they would like to know why model m is proposed instead of model m' where model m' might represent their mental model. Therefore, we adopt tradeoffs-focused explanation templates following Sukkerd et al. (2020). The template takes the form of "(the model) can improve/worsen these performance measures by these amounts, by taking these actions". The proposed explanation templates have been validated by a human subjects experiment. The results show that the tradeoffs-focused explanation significantly improves the users' confidence and trust in the result. Explaining like a dialogue supports interaction and brings all parties closer to modeling (Hoffmann and Magazzeni 2019; Chakraborti et al. 2020; Sukkerd et al. 2020).

3.2.1 Model-Based Explanation

The model-based explanation is used to answer the basic question in the modeling process: "*what are the differences between the currently proposed model A (m_A) and the model B (m_B) under scenario s?*" The alternative model m_B could be another candidate model design or the stakeholder's mental model.

The explanation contains two parts. The first part explains the differences in model configuration, which is capable to capture and express the options and differences explicitly. The second part explains the tradeoffs between a set of *selected* performance measures, denoted by Π. Since people may be uncertain about the objectives at first (Calder et al. 2018), the set can be refined iteratively to help modelers and other stakeholders to discover the preferences together. We propose a predefined natural-language template for generating a model-based explanation:

"*Under scenario s, compared with model B, model A includes options [$\Delta(m_A, m_B)$], but excludes options [$\Delta(m_B, m_A)$]. Such configuration differences improve [these performance measures $\{i : \pi_i(m_A, s) > \pi_i(m_B, s), \pi_i \in \Pi\}$] by [these amounts $\{\pi_i(m_A, s) - \pi_i(m_B, s) : \pi_i(m_A, s) > \pi_i(m_B, s), \pi_i \in \Pi\}$], but also worsen [other performance measures $\{i : \pi_i(m_A, s) < \pi_i(m_B, s), \pi_i \in \Pi\}$] by [these amounts $\{\pi_i(m_B, s) - \pi_i(m_A, s) : \pi_i(m_A, s) < \pi_i(m_B, s), \pi_i \in \Pi\}$].*"

3.2.2 Scenario-Based Explanation

Given a fixed model design, *scenario-based explanation* is used to answer the following question: "*what are the differences that the currently proposed model m performs under scenario 1 (s_1) and scenario 2 (s_2)?*" Different models might be preferred under different business scenarios facing the company. Naturally, users would like to explore the model performance under future scenarios (Calder et al. 2018) and determine whether the currently proposed model is acceptable or not. We propose the following predefined natural-language template for generating a scenario-based explanation:

"*Compared with scenario 2, scenario 1 leads to improvements in [these performance measures* $\{i : \pi_i(m,s_1) > \pi_i(m,s_2), \pi_i \in \Pi\}$*] by [these amounts* $\{\pi_i(m,s_1) - \pi_i(m,s_2) : \pi_i(m,s_1) > \pi_i(m,s_2), \pi_i \in \Pi\}$*], but also worsens [other performance measures* $\{i : \pi_i(m,s_1) < \pi_i(m,s_2), \pi_i \in \Pi\}$*] by [these amounts* $\{\pi_i(m,s_2) - \pi_i(m,s_1) : \pi_i(m,s_1) < \pi_i(m,s_2), \pi_i \in \Pi\}$*].*"

3.2.3 Goal-Oriented Explanation: Embedding Optimization Problems to Answer Open Questions

The above two types of tradeoff-focused explanations are capable to complete simple model comparisons. However, if the modeling task is goal-oriented and the performance of the proposed model does not live up to the stakeholder's expectation, the model-based and scenario-based explanations are not sufficient. Therefore, the goal-oriented explanation is supposed to answer the following open question: "*Under scenario s, if we want to improve the performance measure i of the currently proposed model m by α%, which options should we include/exclude in/from the model?*"

To explain the above open question, we formulate an optimization problem to find a new model m'. According to Camm (2018), we should keep persistence since dramatic changes may lead to managerial resistance to change. (Miller 2019) also confirms that explanations should be selected. That's to say, the information presented should be as succinct as possible. For example, there might be several alternative options to answer an open question, but we don's have to provide all the alternatives exhaustively in the explanation. Therefore, the objective is to find a new model that achieves the improvement with minimal change. The degree of change is measured by the number of inconsistent options, including both addition and deletion. Conclusively, we have to solve the following optimization problem:

$$\min_{m' \in \mathscr{M}} \quad |\Delta(m,m')| + |\Delta(m',m)| \tag{1}$$

$$\text{s.t.} \quad \frac{\pi_i(m',s) - \pi_i(m,s)}{|\pi_i(m,s)|} \geq \alpha \tag{2}$$

The first part in the objective function (1) counts the number of deleted options in the new model m' and the second part counts the number of added options in the new model m'. The performance improvement required by the open question is modeled by constraint (2). We expect three possible outcomes of the optimization problem. First, the problem is infeasible. Explainable modeling is supposed to provide the following explanation:

"*[Explanation when infeasible.] Under scenario s, we cannot find a new model which improves the performance measure i by α%. Please decrease α% and try again.*"

Second, the problem is feasible and the optimal solution is unique. To provide more information about the alternative model m', a model-based explanation of model m and m' can be attached to the following explanation:

"*[Explanation with unique solution.] Under scenario s, to improve the performance measure i of the currently proposed model m by α%, we should include [these options* $\Delta(m',m)$*] and exclude [these options* $\Delta(m,m')$*].*"

Third, the problem is feasible and the optimal solution is not unique. Therefore, we have a set of model, denoted by $\mathscr{M}' = \{m'_1, m'_2, \cdots\}$. We can rank the new models by some predefined rules, such as

minimal addition and minimal deletion. Generally, to provide more information about the alternatives, pairwise model-based comparisons can be attached to the following explanation:

"*[Explanation with multiple solution.] Under scenario s, to improve the performance measure i of the currently proposed model m by α%, we propose several alternatives: (1) including [these options $\Delta(m'_1, m)$] and excluding [these options $\Delta(m, m'_1)$]; (2) including [these options $\Delta(m'_2, m)$] and excluding [these options $\Delta(m, m'_2)$] .....*"

4 MEASURING THE VALUE OF EXPLAINABLE MODELING

To analyze the value of explainable modeling, we introduce three explainability scores. Measuring the value of explanation is important to overcome the challenge of building trust (Shin 2021). An important motivation of explainable modeling is to express the model and key performance tradeoffs explicitly to avoid misunderstanding. Therefore, the value of explainability scores should be a function of the differences between two contrasted models or scenarios. In other words, the explainability score represents the value of eliminating the misunderstanding of differences that would have occurred without clear explanations.

The functional form of explainability scores follows the explicability score proposed by Kulkarni et al. (2019). They study the interpretability of an AI agent's behavior to a human observer. The explicability score is the negative exponential function of the distance between the agent's behavior and the expected one of the observer. The score takes a value between 0 and 1. The score is higher when the agent's behavior is closer to the expected one.

We make modifications for the explainability scores to be suitable in our setting. First, we specify the distance functions to measure the differences in explainable modeling. Second, we take one minus the negative exponential function to represent the value of explainable modeling in eliminating the misunderstanding. The score still takes value from 0 to 1, but it will be higher when the difference expressed in the explanation is larger.

The difference between models is caused by different options. To quantify the value of explaining the inconsistent option selection between model m and m', we introduce the following explainability scoring function V_{model} in Definition 1. The more different options there are in the two models, the higher the score. If the model m and model m' are the same, the explanation does not provide any valuable information, and the score equals zero.

Definition 1 The value of explaining model configuration difference in a model-based explanation is defined as follows:

$$V_{model}(m, m') = 1 - exp\left(-(|\Delta(m, m')| + |\Delta(m', m)|)\right).$$

The first explainability score V_{model} is not sufficient to consider the value of explaining tradeoffs between performance measures in the modeling process. Let consider a situation where only one option is different in two models but results in a significant difference in some performance measure $\pi_i \in \Pi$. In this situation, the explanation expressing performance measures tradeoffs explicitly plays a vital role in the two-way communication between modelers and stakeholders and deserves a higher score. Therefore, we propose the second explainability scoring function $V_{tradeoff}$ in Definition 2.

Definition 2 The value of explaining performance measure tradeoffs in a model-based explanation is defined as follows:

$$V_{tradeoff}(m, m') = \max_{\pi_i \in \Pi, m \neq m'} \left\{ 1 - exp\left(- \frac{|\pi_i(m, s) - \pi_i(m', s)| / |\pi_i(m, s)|}{|\Delta(m, m')| + |\Delta(m', m)|} \right) \right\}.$$

Various business scenarios also lead to different model performances. The explanation is also valuable if it helps stakeholders and modelers figure out how the model performance varies under diverse business scenarios. To quantify the value of scenario-based explanation, we develop the third explainability scoring function $V_{scenario}$ in Definition 3.

Definition 3 The value of scenario-based explanation is defined as follows:

$$V_{scenario}(s,s') = \max_{\pi_i \in \Pi} \left\{ 1 - exp\left(-\frac{|\pi_i(m,s) - \pi_i(m,s')|}{|\pi_i(m,s)|} \right) \right\}.$$

The above three explainability scores complete the framework of explainable modeling. In the next section, we investigate how explainable modeling is applied in building a DT factory.

5 CASE STUDY

5.1 Problem Formulation

In this subsection, we specify the underlying concepts in this case, including options, scenarios, and performance measures. Each option represents a set of constraints that can be included in the model of the DT factory. Options include the practical constraints from demand forecast, delivery plan, to machine capacity. We summarize the options, associated data sources and current integration approaches in Table 2. Options labeled by "(Y)" are mandatory in the model. We specify the notations for other alternative options, such as o_{batch}, o_{change}, o_{limit} and o_{group}. Including options that are manually inputted into the systems will cause additional costs due to the requirement of digitization. Due to the seasonal demand, we consider two business scenarios: (1) high-variety, low-volume; (2) low-variety, high-volume. Let s_{var} and s_{vol} denote the two scenarios separately. The second scenario is common during periods of peak demand.

We consider performance measures from three perspectives. First, the scheduling problem is a multi-objective problem with the following solution quality measures: on-time delivery rate π_{del}, production compliance rate π_{comp}, number of total changeover time π_{change}, machine utilization rate π_{util}, production balancing rate in three workshop areas (A, B, and C), $\pi_{bal-A}, \pi_{bal-B}, \pi_{bal-C}$. Second, the model complexity is measured by π_{run} which is the running time (second) to solve the production planning and scheduling problem. The multi-objective scheduling problem is formulated as a mixed-integer programming (MIP) model. It is first relaxed and solved by a standard solver. After that, we develop a heuristic to fix and improve the solution. The constraints in the MIP model and functional modules in the heuristic are both highly configurable. Third, data integration cost is measured by the number of data input that is currently integrated manually, denoted as π_{data}. In conclusion, there are nine performance measures in the modeling process. We expect a higher value of the following performance measures: π_{del}, π_{comp} and π_{util}; other performance measures are expected to be as small as possible.

Table 2: Option, data source, and current integration approach.

Option	Data source	Current integration approach
Demand forecast (Y)	ERP	Retrieve from ERP
Delivery plan (Y)	ERP	Retrieve from ERP
Batch volume (o_{batch})	ERP	Retrieve from ERP
Machine changeover time (o_{change})	MES	Retrieve from MES
Machine capacity (Y)	Current scheduling system	Manual input
Machine capability (Y)	Current scheduling system	Manual input
Total production quantity limit (o_{limit})	Current scheduling system	Manual input
Working calendar (Y)	Current scheduling system	Manual input
Machine maintenance plan (Y)	Emails	Manual input
Machine group (o_{group})	Not available online	Manual input

5.2 Explanations and Analysis

According to Table 2, four alternative options lead to 16 candidate models. Each model is evaluated by nine performance measures under two business scenarios (a total of 288 performance measures). Without the

explainable modeling framework, we, as the modelers, and domain experts always identified the problems with different languages. Due to a lack of explicit expression, it was time-consuming to reach a consistent understanding of how options and scenarios interacted to influence model performances and which option was required.

5.2.1 Model-Based Explanation and Analysis

We first illustrate the model-based explanations by comparing the mental model of the end-user and the most complex model. Let m_0 be the end-user's mental model which is used to schedule the production manually. We have $m_0(o_{batch}) = m_0(o_{change}) = m_0(o_{limit}) = m_0(o_{gourp}) = 0$. The end-user is unable to consider the batch volume, sequence-dependent changeover time, production quantity limit, and machine group constraints. Let m_{full} be the model where $m_{full}(o_{batch}) = m_{full}(o_{change}) = m_{full}(o_{limit}) = m_{full}(o_{gourp}) = 1$. Instead of discussing the detailed solution to the end-users, the following model-based explanation directly illustrates how the two models perform differently under the two business scenarios.

- Model-based explanation 1: *under scenario s_{var}, compared with model m_0, model m_{full} includes options [$o_{batch}, o_{change}, o_{limit}, o_{gourp}$]. Such configuration differences improve the [production compliance rate π_{comp}] by [increasing* 1.57%*], [the number of total changeover time π_{change} by [decreasing* 10 *times] , [machine utilization rate π_{util}] by [increasing* 6.92%*], and [production balancing rate in workshop area A π_{bal-A}] by [decreasing* 0.22*]; but also worsens [the running time π_{run} by [increasing* 250 *seconds], [data integration cost π_{data}] by [increasing additional* 486 *data points collection per month], and [production balancing rate in workshop area B π_{bal-B}] by [increasing* 0.55*].*
- Model-based explanation 2: *under scenario s_{vol}, compared with model m_0, model m_{full} includes options [$o_{batch}, o_{change}, o_{limit}, o_{gourp}$]. Such configuration differences improve the [on time delivery rate π_{del}] by [increasing* 0.12%*], [production compliance rate π_{comp}] by [increasing* 1.05%*], [the number of total changeover time π_{change} by [decreasing* 26 *times] , [machine utilization rate π_{util}] by [increasing* 10.04%*], [production balancing rate in workshop area A π_{bal-A}] by [decreasing* 0.99*], [production balancing rate in workshop area B π_{bal-B}] by [decreasing* 6.22*], and [production balancing rate in workshop area C π_{bal-C}] by decreasing* 1.41*]; but also worsens [the running time π_{run} by [increasing* 90 *seconds], [data integration cost π_{data}] by [increasing additional* 486 *data points collection per month].*

The model-based explanations show that the intelligent scheduling system can significantly improve the solution quality compared with the manually generated solution. However, the end-users and developers would like to know how the model m_{full} can be modified to further improve performance measures of their interests. For example, the changeover of a production line is a highly time-consuming activity and the end-users want to reduce the number of total changeover time to reduce the setup time and the workload of workers. We show how explainable modeling is used to provide goal-oriented explanation.

- Question: *under scenario s_{var}, if we want to improve the number of total changeover time of the currently proposed model m_{full} by* 5%*, which options should we include/exclude in/from the model?*
- Goal-oriented explanation 1: *under scenario s_{var}, to improve the number of total changeover time of the currently proposed model m_{full} by* 5%*, we propose several alternatives: (1) excluding [the option of batch volume o_{batch}]; (2) excluding [the option of total production limit o_{limit}].*

In order to answer the question, we specify the constraint (2) introduced in Section 3.2.3 as $\pi_{change}(m_{full}, s_{var}) - \pi_{change}(m', s_{var}) \geq 5\% \cdot \pi_{change}(m_{full}, s_{var})$ while minimizing the number of different options. In this case, we find two alternative models with only one different option with model m_{full}. One (m_1) excludes the option of batch volume with $\pi_{change}(m_1, s_{var}) = 14$ and the other (m_2) excludes the option of total production

quantity limit with $\pi_{change}(m_2, s_{var}) = 15$. In order to reduce the number of changeover time, the end-users may prefer model m_1 than m_2. To determine whether to exclude option o_{batch} and use model m_1 instead of model m_{full}, a model-based explanation between model m_1 and model m_{full} can be used again to facilitate the discussion. The model-based explanation below highlights the tradeoff information of using model m_1 instead of m_{full}.

- Model-based explanation 3: *under scenario s_{var}, compared with model m_{full}, model m_1 exclude option [o_{batch}]. Such configuration differences improve the [production compliance rate π_{comp}] by [increasing* 1.49%*], [the number of total changeover time π_{change} by [decreasing* 2 *times] , [the running time π_{run} by [decreasing* 104 *seconds], and [data integration cost π_{data}] by [decreasing additional* 62 *data points collection per month]; but also worsens [machine utilization rate π_{util}] by [decreasing* 6.2%*], [production balancing rate in workshop area A π_{bal-A}] by [increasing* 0.23*], and [production balancing rate in workshop area B π_{bal-B}] by [increasing* 0.89*].*

Finally, we use the explainability scores introduced in Section 4 to illustrate the value of model-based explanations. We compute the explainability scores V_{model} and $V_{tradeoff}$ based on model-based explanations 1 and 3 since both of these two explanations are under the business scenario s_{var}. As shown in Equation 3, the model-based explanation 1 has a higher explainability score based on Definition 1 because explanation 1 clarifies more option differences than explanation 3. Let $\Pi = \{\pi_{change}\}$. Explanation 3 has a higher value from the perspective of explaining tradeoffs since it expresses clearly how a slight option modification impacts the performance measure of interest, as shown in Equation 4. However, a limitation of our explainability scores is that there is no comparison between explainability scores V_{model} and $V_{tradeoff}$. A more consistent scoring function to measure the explainability is left for future research.

$$V_{model}(m_0, m_{full}) = 1 - exp(-4) = 0.98 > V_{model}(m_{full}, m_1) = 1 - exp(-1) = 0.63. \quad (3)$$

$$V_{tradeoff}(m_0, m_{full}) = 1 - e^{-\frac{|26-16|/26}{4}} = 0.09 < V_{tradeoff}(m_{full}, m_1) = 1 - e^{-\frac{|16-14|/16}{1}} = 0.11. \quad (4)$$

5.2.2 Scenario-Based Explanation and Analysis

Assumed that model m_1 is preferred due to the less changeover time and additional data integration cost. It remains unclear how the model m_1 performs under the other business scenario s_{vol}. Therefore, we present the following scenario-based explanation to inform the end-users of the scenario's impact on performance measures. For example, the production balancing rates in workshop areas B and C are significantly reduced because no product needs to be produced in these two areas (low-variety). Meanwhile, the production balancing rate in workshop area A is increased due to the feature of high volume.

- Scenario-based explanation 1: *compared with scenario s_{var}, scenario s_{vol} leads to improvements in [running time π_{run}] by [decreasing* 79 *seconds], [machine utilization rate π_{util}] by [increasing* 10.23%*], and [production balancing rate in workshop area B π_{bal-B}] by [decreasing* 4.47*]; but also worsens [production compliance rate π_{comp}] by [decreasing* 2.73%*], [the number of changeover time] by [increasing* 3 *times], and [production balancing rate in workshop area A π_{bal-A}] by [increasing* 1.62*].*

Let $\Pi = \{\pi_{del}, \pi_{comp}, \pi_{change}, \pi_{util}, \pi_{bal-A}, \pi_{bal-B}, \pi_{bal-C}\}$ that includes all the solution quality related performance measures. We calculate the value of scenario-based explanation 1 based on Definition 3 in the equation below:

$$V_{scenario}(s_{var}, s_{vol}) = \max_{\pi_i \in \Pi} \left\{ 1 - exp\left(-\frac{|\pi_i(m_1, s_{var}) - \pi_i(m_1, s_{vol})|}{|\pi_i(m_1, s_{var})|} \right) \right\} = 0.69.$$

It is shown that the largest impact of scenario s_{vol} on performance measures is on the production balancing rate in workshop area A. If the end-users would like to explore how the model m_1 can be modified to improve the production balancing rate π_{bal-A}, another round of open question and model-based explanations can be conducted as shown in Section 5.2.1.

In summary, we illustrate how explainable modeling is applied in the modeling process of building a DT factory. The automatically generated explanations explicitly express the tradeoffs between different models and scenarios. The open question is helpful to increase stakeholders' engagement in exploring the best model under various scenarios.

6 CONCLUSIONS

We propose an explaining modeling framework to help modelers and stakeholders collaborate closely in the modeling process. It is applicable in various domains. In particular, we present a case study in a DT factory project. We suggest three directions for future research. First, more explanation generation approaches should be explored. For example, stakeholders may ask about the interaction between different options and raise other open questions. Second, we have not discussed the analytics and algorithms to generate explanations efficiently. Third, a more consistent explainability score is required to analyze the quality of explanation.

REFERENCES

Bertsimas, D., and A. King. 2016. "OR Forum – An Algorithmic Approach to Linear Regression". *Operations Research* 64(1):2–16.

Calder, M., C. Craig, D. Culley, R. de Cani, C. A. Donnelly, R. Douglas, B. Edmonds, J. Gascoigne, N. Gilbert, C. Hargrove et al. 2018. "Computational Modelling for Decision-Making: Where, Why, What, Who and How". *Royal Society Open Science* 5(6):172096–1–172096–15.

Camm, J. D. 2018. "How to Influence and Improve Decisions through Optimization Models". In *Recent Advances in Optimization and Modeling of Contemporary Problems*, 1–19. The Institute for Operations Research and the Management Sciences.

Cerić, A. 2015. "Trust in Construction Projects: Literature Analysis Using Keywords". *Organization, Technology and Management in Construction: an International Journal* 7(1):1179–1185.

Chakraborti, T., S. Sreedharan, and S. Kambhampati. 2020. "The Emerging Landscape of Explainable Automated Planning & Decision Making". In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 4803–4811. January, Yokohama, Japan: International Joint Conferences on Artificial Intelligence Organization.

Digital Twin Consortium 2020. "Digital Twin Consortium Defines Digital Twin". https://blog.digitaltwinconsortium.org/2020/12/digital-twin-consortium-defines-digital-twin.html, accessed 4th Apr.

Du, J., D. Zhao, R. R. Issa, and N. Singh. 2020. "BIM for Improved Project Communication Networks: Empirical Evidence from Email Logs". *Journal of Computing in Civil Engineering* 34(5):04020027–1–04020027–14.

Du, J., Q. Zhu, Y. Shi, Q. Wang, Y. Lin, and D. Zhao. 2020. "Cognition Digital Twins for Personalized Information Systems of Smart Cities: Proof of Concept". *Journal of Management in Engineering* 36(2):04019052–1–04019052–17.

Fuller, A., Z. Fan, C. Day, and C. Barlow. 2020. "Digital Twin: Enabling Technologies, Challenges and Open Research". *IEEE Access* 8:108952–108971.

Harper, A., N. Mustafee, and M. Yearworth. 2021. "Facets of Trust in Simulation Studies". *European Journal of Operational Research* 289(1):197–213.

Hoffmann, J., and D. Magazzeni. 2019. *Explainable AI Planning (XAIP): Overview and the Case of Contrastive Explanation*, 277–282. Cham: Springer International Publishing.

Honti, G., G. Dörgö, and J. Abonyi. 2019. "Review and Structural Analysis of System Dynamics Models in Sustainability Science". *Journal of Cleaner Production* 240:118015–1–118015–25.

Jahangirian, M., S. Borsci, S. G. S. Shah, and S. J. Taylor. 2015. "Causal Factors of Low Stakeholder Engagement: A Survey of Expert Opinions in the Context of Healthcare Simulation Projects". *Simulation* 91(6):511–526.

Jones, D., C. Snider, A. Nassehi, J. Yon, and B. Hicks. 2020. "Characterising the Digital Twin: A Systematic Literature Review". *CIRP Journal of Manufacturing Science and Technology* 29:36–52.

Kolesnikov, S., N. Siegmund, C. Kästner, A. Grebhahn, and S. Apel. 2019. "Tradeoffs in Modeling Performance of Highly Configurable Software Systems". *Software & Systems Modeling* 18(3):2265–2283.

Kulkarni, A., S. Sreedharan, S. Keren, T. Chakraborti, D. E. Smith, and S. Kambhampati. 2019. "Design for Interpretability". In *Proceedings of the Second International Workshop on Explainable AI Planning (XAIP)*, 1–5. July 12th, Berkeley, CA, USA: International Conference on Automated Planning and Scheduling, Inc.

Levasseur, R. E. 2010. "People Skills: Ensuring Project Success—A Change Management Perspective". *Interfaces* 40(2):159–162.

Lu, Q., X. Xie, A. K. Parlikad, and J. M. Schooling. 2020. "Digital Twin-Enabled Anomaly Detection for Built Asset Monitoring in Operation and Maintenance". *Automation in Construction* 118:103277–1–103277–16.

Lupeikiene, A., G. Dzemyda, F. Kiss, and A. Caplinskas. 2014. "Advanced Planning and Scheduling Systems: Modeling and Implementation Challenges". *Informatica* 25(4):581–616.

Miller, T. 2019. "Explanation in Artificial Intelligence: Insights from the Social Sciences". *Artificial Intelligence* 267:1–38.

Norman, D. A. 1983. "Some Bbservations on Mental Models". *Mental Models* 7(112):7–14.

Norman, D. A. 1988. *The Psychology of Everyday Things.* New York: Basic books.

Pan, Y., and L. Zhang. 2021. "A BIM-Data Mining Integrated Digital Twin Framework for Advanced Project Management". *Automation in Construction* 124:103564–1–103564–15.

Robinson, S. 2015. "A Tutorial on Conceptual Modeling for Simulation". In *Proceedings of the 2015 Winter Simulation Conference*, 1820–1834. Dec 6th-9th, Huntington Beach, CA, USA: Institute of Electrical and Electronics Engineers, Inc.

Shao, G., S. Jain, C. Laroque, L. H. Lee, P. Lendermann, and O. Rose. 2019. "Digital Twin for Smart Manufacturing: The Simulation Aspect". In *Proceedings of the 2019 Winter Simulation Conference*, 2085–2098. December 8th-12th, National Harbor Maryland, USA: Institute of Electrical and Electronics Engineers, Inc.

Shin, D. 2021. "The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI". *International Journal of Human-Computer Studies* 146:102551–1–102551–10.

Siegmund, N., A. Grebhahn, S. Apel, and C. Kästner. 2015. "Performance-Influence Models for Highly Configurable Systems". In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, 284–294. Aug 30th-Sep 4th, Bergamo, Italy: Association for Computing Machinery.

Stark, R., and T. Damerau. 2019. *Digital Twin*, 1–8. Berlin, Heidelberg: Springer Berlin Heidelberg.

Sukkerd, R., R. Simmons, and D. Garlan. 2020. "Tradeoff-Focused Contrastive Explanation for MDP Planningg". In *Proceedings of the 2020 IEEE International Conference on Robot and Human Interactive Communication*, 1041–1048. Aug 31th-Sep 4th, Naples, Italy: Institute of Electrical and Electronics Engineers, Inc.

Taillandier, P., A. Grignard, N. Marilleau, D. Philippon, Q.-N. Huynh, B. Gaudou, and A. Drogoul. 2019. "Participatory Modeling and Simulation with the GAMA Platform". *Journal of Artificial Societies and Social Simulation* 22(2):1–20.

Techtarget 2008. "How Do APS and ERP Fit Together?". https://searcherp.techtarget.com/answer/How-do-APS-and-ERP-fit-together, accessed 4th, Apr.

UNSW Newsroom 2020. "The Digital Technology Set to Transform the Construction Industry". https://newsroom.unsw.edu.au/news/art-architecture-design/digital-technology-set-transform-construction-industry, accessed 4th, Apr.

Wang, K., W. Xie, B. Wang, J. Pei, W. Wu, M. Baker, and Q. Zhou. 2020. "Simulation-Based Digital Twin Development for Blockchain Enabled End-to-End Industrial Hemp Supply Chain Risk Management". In *Proceedings of the 2020 Winter Simulation Conference*, 3200–3211. December 14th-18th, Orlando, Florida, USA: Institute of Electrical and Electronics Engineers, Inc.

Weisel, E. W., M. D. Petty, and R. R. Mielke. 2003. "Validity of Models and Classes of Models in Semantic Composability". In *Proceedings of the Fall 2003 Simulation Interoperability Workshop*, Volume 9, 68. September 14th-19th, Orlando, Florida, USA: Citeseer.

Yilmaz, L., and B. Liu. 2020. "Model credibility revisited: Concepts and considerations for appropriate trust". *Journal of Simulation* 0(0):1–14.

AUTHOR BIOGRAPHIES

LU WANG is Ph.D. candidate in the Department of Industrial Engineering at Tsinghua University. Her research interest includes DT and production planning and scheduling. Her email address is wanglu18@mails.tsinghua.edu.cn.

TIANHU DENG is an associate professor in the Department of Industrial Engineering at Tsinghua University. He received his Ph.D. in Industrial Engineering and Operations Research from the University of California Berkeley. He has done research in DT and supply chain management. His email address is deng13@tsinghua.edu.cn.

ZEYU ZHENG is an assistant professor in the Department of Industrial Engineering & Operations Research at the University of California Berkeley. He received his Ph.D. in Management Science and Engineering from Stanford University. He has done research in simulation, stochastic modeling, data analytics, and statistical learning. His email address is zyzheng@berkeley.edu.

ZUO-JUN MAX SHEN is a professor in the Department of Industrial Engineering & Operations Research at the University of California Berkeley. He received his Ph.D. in Industrial Engineering and Management Sciences from Northwestern University. He has done research in integrated supply chain management and optimization algorithms. His email address is maxshen@berkeley.edu.