Verification and Validation of D2FD Method

Yan Wang, Mamadou Kaba Traoré, Grégory Zacharewicz, David Chen

Verification and Validation of D2FD Method

Yan Wang
University of Bordeaux - Bat A31
Laboratory IMS UMR CNRS 5218
351 Cours de la Libération, 33405 Talence cedex, FRANCE
yan.wang@u-bordeaux.fr

Grégory Zacharewicz
University of Bordeaux - Bat A31
Laboratory IMS UMR CNRS 5218
351 Cours de la Libération, 33405 Talence cedex, FRANCE
gregory.zacharewicz@u-bordeaux.fr

Mamadou Kaba Traoré
University of Blaise Pascal
Laboratory LIMOS UMR CNRS 6158
1 rue de la Chebarde, 63178 Aubiere cedex, FRANCE
traore@isima.fr

David Chen
University of Bordeaux - Bat A31
Laboratory IMS UMR CNRS 5218
351 Cours de la Libération, 33405 Talence cedex, FRANCE
david.chen@u-bordeaux.fr

***Abstract*—D2FD (Data to Fuzzy-DEVS) method provides a solution for the problem of system inference. This method is well designed and implemented as an available and dedicated plug-in within the process mining framework (ProM). This plug-in is also integrated with the simulation tool SimStudio. However, the last step of the process of inferring models and simulations, which is verification and validation, is missing. This paper proposes a new paradigm of verification and validation in system inference. The case study uses the method of comparing with other models as the main validation technique. Based on the same data source from the Dutch Employee Insurance Agency, it attempts to compare the previous results with other studies.**

Keywords—system inference; modeling and simulation; D2FD method; verification; validation; ProM; SimStudio

I. Introduction (*Heading 1*)

System theory [1] provides some methods to describe systems. The knowledge of the system can be organized in a 4-level hierarchy [2]. It consists the source level, the data level, the generative level and the structure level. Moving between these levels, there are three systems problems: in system analysis, we have an existing system and we are trying to review or create source; in system design, the system doesn't exist yet and we are investigating the alternative structures for a new system; in system inference, the system exists and we are trying to generate the structure. Among them, system inference is a big challenge of system modeling.

D2FD method provides a method to deal with system inference and mine a discrete event simulation model. This method contains three parts. In the first part, it provides a new method to transform the event data to the event logs. The event data is observed from real world and the event logs is recorded based on XES (Extensible Event Stream) standard [3]. In the second part, a method of transforming event logs into transition systems [4] is included. This method is coming from process mining [3]. In the third part, a new method of transforming transition system into Fuzzy-DEVS (Discrete Event System Specification) is proposed. Fuzzy-DEVS model [5] not only provides a general framework to represent dynamic systems, but also represents more information including the modularity, frequency and timing aspects.

The original process of developing models and simulations has three steps. A conceptual model is designed by a subject matter expert (SME) from careful consideration of a problem and its domain. Then, it is realized via a source code simulation through the implementation of interfaces, data structures and algorithms. Finally, the output of the simulation for a set of test cases is validated against historical data or other trusted sources [6]. The process of inferring models and simulations is a little different in the first steps. The first step is that an inference model is mined from the data or the problem in the reality using some mining techniques. However, the last two steps are the same. The inferred model is used for simulation, verification and validation. Model verification is often defined as "ensuring that the computer program of the computerized model and its implementation are correct". Model validation is usually defined to mean "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" [7].

Sargent [8] proposes a paradigm that relates verification and validation to the model development process. This paradigm shows a Real World and a Simulation World. However, this paradigm is based on the process of developing models and simulations. In this paper, we propose a new paradigm of the process of inferring models and simulations. Based on this paradigm, we explain the position of D2FD method and supply this method with one validation method. The validation method is selected from all the validation techniques [8]. At last, the case study chooses the Dutch Employee Insurance Agency [9] as the data source. We compare the previous results [10] and the other studies [11] for its validation.

This paper is organized as follows. Section 2 introduces some related studies. Section 3 introduces the new paradigm. Section 4 briefly shows the D2FD method and locates it in the paradigm. Section 5 explains the main process of the selected validation method. The case study is shown in Section 6. The final conclusion is in Section 7.

II. RELATED STUDIES

The design of a model that appeared complete and robust can become incoherent, incomplete and potentially invalid during simulation implementation [8]. Exploring alternative models and alternative modeling questions becomes impossible because SMEs are unable to identify the respective modifications that need to be made to the simulation [12]. It is important to make verification and validation.

A lot of techniques for building valid and credible simulation models are proposed [13][14][15]. Gore et al. [16] provides the approach used in practice with a formal specification languages. They force attention to mathematical details. However, there are no related studies which focus on the process of inferring models and simulations.

III. NEW PARADIGM

The proposed paradigm is based on the paradigm in [8] and the process of inferring models and simulation. It is shown in Figure 1.There are a Real World and a Simulation World. In the Real World, there exist some systems or problems. System theories describe the characteristics of the system and its behavior. System data is obtained by conducting experiments on the system. System theories are developed by abstracting what has been observed from the system and by hypothesizing from the system data and results. System theories are validated by performing theory validation. Theory validation involves the comparison of system theories against system data and results over the domain the theory is applicable for to determine if there is agreement. This process requires numerous experiments to be conducted on the real system.

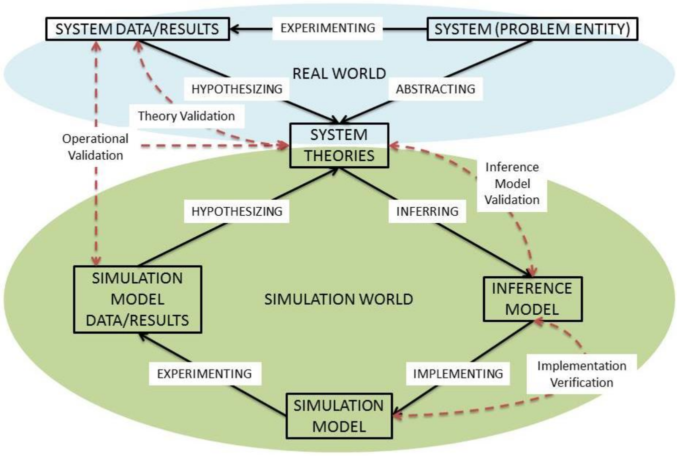

Figure 1: The structure of the inferring process with verification and validation.

We now discuss about the Simulation World which contains four processes. Simulation model is always developed based on objectives. The inference model is the mathematical discipline to represent the system for the objectives of a particular study. The simulation model is the inference model running on the computer system like ProM. The simulation model data and results are the data and results from experiments conducted on the simulation model. The inference model is mined from the data or problem in the Real World. If it is a Fuzzy-DEVS model, the method of inferring is based on D2FD method. Then this model is implemented through a plugin [17] available in ProM and experimented in SimStudio [18]. Inference model validation is defined as determining that the theories and assumptions underlying the inference model are consistent with those in the system theories. Inference verification is defined as assuring that the software design and the specification for programming and implementing the inference model on the specified computer system is satisfactory. Operational validation is defined as determining that the model's output behavior has sufficient accuracy for the model's intended purpose over the domain of the model's intended applicability.

IV. D2FD METHOD

The D2FD method has three major stages, as presented in Figure 2: (1) from event data to event logs; (2) from event logs to transition system; (3) from transition system to Fuzzy-DEVS model. Here we introduce the general structure of D2FD method.

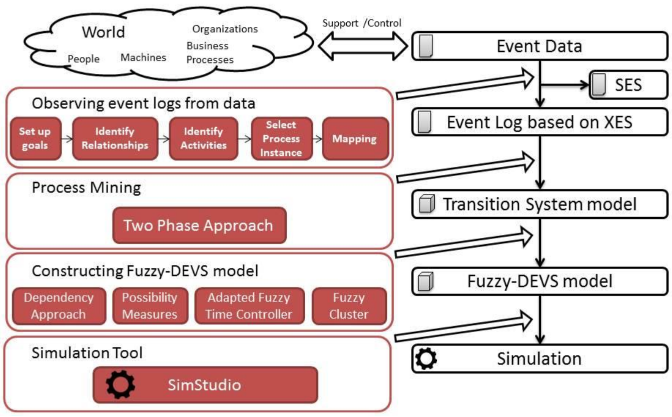

Figure 2: The structure of D2FD method.

A. *From Event Data to Event Logs*

The data is observed from the world in the data level of Figure 1. However, not every data is available for D2FD method. The event data are the starting point and they are selected through several guidelines. The guidelines include the twelve guidelines from Van der Aalst [3] and four more guidelines. When the event data is observed, five steps are proposed to transform event data to event logs as shown in Figure 2. First, we need to set up the goal by the interview of business people. The goal can be the business problem which we are going to solve. Second, System Entity Structure (SES) [19] defines a ontological framework in the systems theory. We construct SES from event data in order to discover the relationships between the activities and refine the event data. Third, we identify the activities as well as the modularity. The activities can be identified as public activity and private activity. If some activities have a strong relationship with some activities in other documents, we can define their children activities as private activities. Conversely, public activities. Fourth, process instance relates to the object that you will follow throughout the process. As we get the hierarchy of SES structure, we select the interesting level according to the goals. This interesting level is related the one of the attributes and contains several activities. Fifth, the activities of interesting attribute, time, case and instance are transformed into the parameters of the event logs.

B. Form Event Logs to Transition System

The method of transforming from event logs to transition system is coming from Two Phase Approach. State and transition can be mined from event logs. Let S^T be the state and T be the transition. In this method, there are different methods in different dimensions to capture states, which lead to different kinds of transition systems. For example, past or future; set abstraction or multi-set abstraction; k-tail method. The constructed transition system is considered as a low-level process model.

C. From Transition System to Fuzzy-DEVS Model

Fuzzy-DEVS model is based on the Fuzzy-DEVS formalism which applies fuzzy sets theory into the functions of DEVS formalism. It consists of Fuzzy-DEVS atomic model and Fuzzy-DEVS coupled model. A fuzzy atomic model $\tilde{M}$ is characterized by:

$$\tilde{M} = \langle X, Y, S, \tilde{\delta}_{int}, \tilde{\delta}_{ext}, \tilde{\lambda}, \tilde{ta} \rangle$$

- X: the set of input values;
- Y: the set of output values;
- S: the set of states;
- $\tilde{\delta}_{int}$: S×S → [0, 1], fuzzy internal transition function;
- $\tilde{\delta}_{ext}$: Q×X×S×S → [0, 1], fuzzy external transition function, Q = {(s, e) | s ∈ S, 0 ≤ e ≤ ta(s)} where ta(s) is the defuzzified value of $\tilde{ta}$;
- $\tilde{\lambda}$: S×Y → [0, 1], fuzzy output function;
- $\tilde{ta}$:S×$\tilde{A}$ → [0, 1], fuzzy time advance function, $\tilde{A}$ = the set of fuzzy linguistic numbers.

A coupled model DN is defined:

$$DN = \langle X, Y, D, EIC, EOC, IC, SELECT \rangle$$

- X: input event sets;
- Y: output events sets;
- D: DEVS components set;
- EIC ⊆ {((N, ip_N), (d, ip_d)) | ip_N∈IPorts, d∈D, ip_d∈$IPorts_d$}.
- EOC ⊆ {((d, op_d), (N, op_N)) | op_N∈OPorts, d∈D, op_d∈$OPorts_d$}.
- IC ⊆ {((a, op_a), (b, ip_b)) | a, b∈D, op_a∈D, ∈$OPorts_a$, ip_b∈$IPorts_b$}.
- SELECT: 2^D→{}→D, tie-breaking selector.

The method of transforming from transition system to Fuzzy-DEVS model is based on the previous work (Wang et al. 2015). In Fuzzy-DEVS atomic model, an improved region-based approach is defined to specify state. Let TS = (S^T, A, T) be a transition system and R ⊆ S^T be a subset of states. P_a is a period time for each activity a ∈ A. R is a region if for each activity a ∈ A and one of the following conditions holds:

- All transition (s_1^T, a, s_2^T) ∈ T enter R, i.e. s_1^T ∉ R and s $_2^T$ ∈ R;
- All transition (s_1^T, a, s_2^T) ∈ T exit R, i.e. s_1^T ∈ R and s_2^T ∉ R;
- All transition (s_1^T, a, s_2^T) ∈ T do not cross R, i.e. s_1^T, s_2^T ∈ R or, s_1^T, s_2^T ∉ R;
- For all the transitions a_1 ∈ T_1, a_2 ∈ T_2, …, a_n ∈ T_n enter R, $P_{a1} \approx P_{a2} \approx \ldots \approx P_{an}$.

Let pa be the private activity and ua be the public activity. According to Fuzzy-DEVS formalism in chapter 3, the transformation follows the rules:

$$R \rightarrow S \tag{1}$$

Where the state of DEVS atomic model s∈S.

$$ua \rightarrow x \cup y \tag{2}$$

Where the input value x∈X and the output value y∈Y.

$$t\tilde{a} = \begin{cases} 0 & \exists s_0 \in S \\ T^F & \\ Infinite & \neg\exists S = S_1^I \end{cases} \tag{3}$$

Where s_0 is the initial state, T^F is the result coming from Adapted Fuzzy Time Controller, S_1^I is the input states of all internal transition.

$$T \rightarrow \tilde{\delta}_{int}$$
$$(s_1^T, ua, s_2^T) \rightarrow (s_1, s_2, \mu_{int}) \tag{4}$$

Where s1 ∈ R1 and s2 ∈ R2, μ_{int} is the result coming from dependency method.

$$\tilde{\lambda} : (y, \mu_{int}) \tag{5}$$

$$T \rightarrow \tilde{\delta}_{ext}$$
$$(s_1^T, pa, s_2^T) \rightarrow (s_1, e, x, s_2, \mu_{ext}) \tag{6}$$

$$\mu(s_i \rightarrow s_j) = \begin{cases} \dfrac{F(s_i \rightarrow s_j) - F(s_j \rightarrow s_i)}{F(s_i \rightarrow s_j) + F(s_j \rightarrow s_i) + 1} \exists i \neq j \\ \dfrac{F(s_i \rightarrow s_j)}{F(s_i \rightarrow s_j) + 1} \exists i = j \end{cases} \tag{7}$$

Where the elapsed time e: 0 ≤ e ≤ $\tilde{ta}$, μ_{ext} is the result coming from the dependency method. The transition system is first divided into regions then transformed into state. The public activities can be the sets of the input or the output values. If the state is the initial state, the fuzzy time is set as 0. If the state has no internal transition, the fuzzy time is set infinite. Otherwise, the fuzzy time is calculated based on Adapted Fuzzy Time Controller (AFTC). Time duration and remaining time are the inputs of AFTC. They first convert into fuzzy sets by membership function and then defuzzify into crisp value. They activate the final fuzzy time based on the rule base. If the transition T contains the public activity, it is transformed into fuzzy internal transition. If the transition T contains the private activity, it is transformed into fuzzy external transition. Both the fuzzy internal transition and fuzzy external transition have the membership function μ. It is measured by Dependency Method. The dependency method is based on Equation 7. The fuzzy output function has the same membership function as fuzzy internal transition.

$$E\tilde{IC} \subseteq \{((N, ip_N), (d, ip_d), (\mu_{EIC}, e)) \mid \mathrm{ip}_N \in IPorts, d \in D, ip_d \in IPorts_d\} \tag{8}$$

$$E\tilde{OC} \subseteq \{((d, op_d), (N, op_N), (\mu_{EOC}, e)) \mid \mathrm{op}_N \in OPorts, d \in D, op_d \in OPorts_d\} \tag{9}$$

$$I\tilde{C} \subseteq \{((a, op_a), (b, ip_b), (\mu_{IC}, e)) \mid \mathrm{a, b} \in D, op_a \in Oports_a, ip_b \in IPorts_b\} \tag{10}$$

In Fuzzy-DEVS coupled model, fuzzy cluster is proposed for integration [10]. Three functions are integrated with membership coefficients. In Equation 8, 9 and 10, μ is the membership coefficient calculated based on Dependency Method and e is the elapsed time. While the time elapses, the membership coefficients may change.

V. VALIDATION TECHNIQUES

The operational validation is a big challenge for D2FD method. It is determining whether the simulation model's output behavior has the accuracy required for the model's intended purpose over the domain of the model's intended applicability. According to [8], a lot of techniques are proposed and collected for the operational validation. In this paper, we choose to use "comparison to other models" as the main technique. This technique focuses on the subject and it is not observable. The simulation model which we construct is compared to other simulation models that have been validated.

VI. CASE STUDY

The case study is from BPI Challenge (2016). It covers the information of the customers and the records of the telephone calls in the category of question and complaints from a Dutch Employee Insurance Agency. The workbook message is contacted by customers through a digital channel (Here the Dutch language is translated into English). For validating the model built by D2FD method, we compare with other simulation models [11] that have been validated. Figure 3 shows the model built from question file by D2FD method. Compared with it, other model which considers the same problem is shown in Figure 4. The process from werkmap to taken is the same but the other processes show different possibilities. The werkmap to taken is not considered as important in Figure 3.

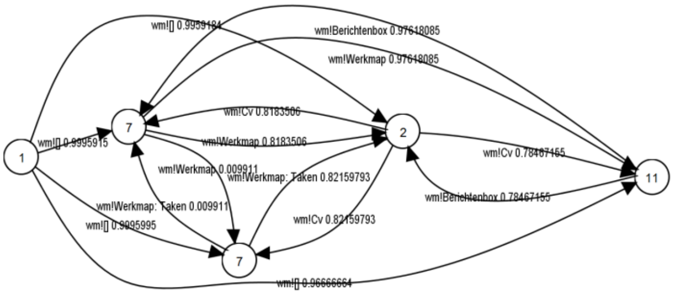

Figure 3: Question Model built by D2FD method.

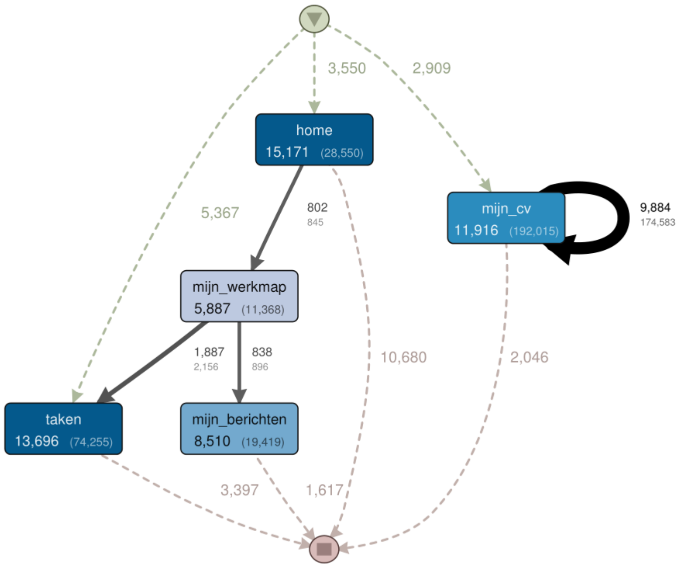

Figure 4: Other model compared with Figure 3.

Another comparison uses the channel file as experiment. Figure 5 shows the model built by D2FD method and Figure 6 shows the validated model. These two models show almost the same process between the channels. However, by using D2FD method, which activity is using channels can be detected. For example, werkmap is using channel 1.

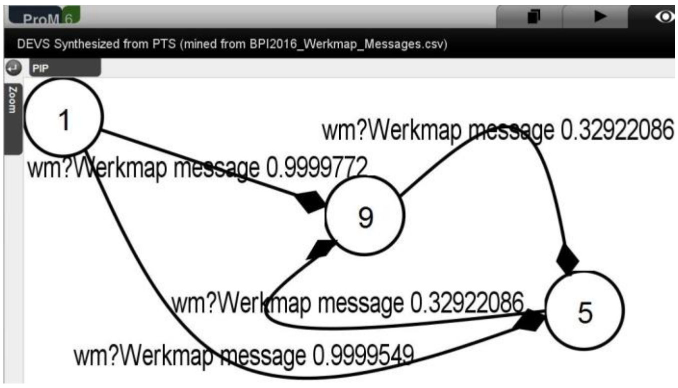

Figure 5: Channel Model built by D2FD method.

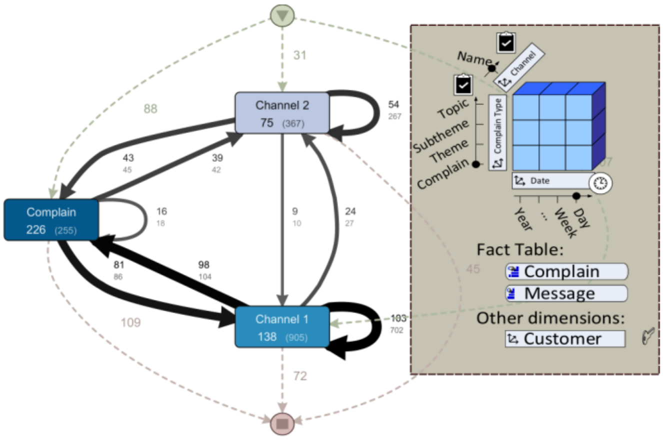

Figure 6: Other model compared with Figure 5.

VII. CONCLUSION

Verification and validation are very important for simulation models. The aim is to prove the coherence of the model, to ensure the correct use of the modeling means and account for the description of the requirements that prevailed in the existence of models. Despite the process of developing models and simulations, a new process of inferring models and simulations is proposed. The paradigm provides a different perspective of modeling and simulation. The method of this paradigm is based on D2FD method. It is available, interoperable, feasible and visible. One of the verification and validation techniques is selected to supply D2FD method. The case study illustrates the validation of the real enterprise problem. In the full paper, the case study will be explained more in detail. We anticipate using more verification and validation techniques for the real problems.

REFERENCES

[1] Simon H.A., The architecture of complexity, Facets of systems science, Springer US, 1991: 457-476.

[2] Klir G., Architecture of systems problem solving, Springer Science & Business Media, 2013.

[3] Van der Aalst W.M.P., *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer Science & Business Media, 2011.

[4] Robert M.K., Formal verification of parallel programs. Communications of the ACM, 19(7), 371-384, 1976.

[5] Kwon Y.W., Park H.C., Jung S. and Kim T.G., Fuzzy-DEVS Formalisme: Concepts, Realization and Application, Proceedings AIS 1996, pp. 227–234, 1996.

[6] Pace, D. K. 2000. “Ideas about simulation conceptual model development”. Johns Hopkins APL technical digest 21 (3): 327–336.

[7] Schlesinger, et al. 1979. Terminology for model credibility. Simulation 32 (3): 103-104.

[8] Sargent R G. Verification and validation of simulation models[C]//Proceedings of the 37th conference on Winter simulation. winter simulation conference, 2005: 130-143.

[9] Business Process Intelligence (BPI) Challenge 2016. https://data.4tu.nl/repository/uuid:360795c8-1dd6-4a5b-a443-185001076eab.

[10] Wang, Y., Zacharewicz, G., Chen D. and Traoré M.K., Use of Fuzzy Clustering for Discrete Event Simulation Model Consruction, IFAC 2017 World Congress, The 20th World Congress of the International Federation of Automatic Control, Toulouse, France. 9-14 July, 2017.

[11] Jalali A., Exploring different aspects of users behaviours in the Dutch autonomous administrative authority through process cubes. Business Process Intelligence (BPI) Challenge 2016.

[12] Law, A. M. 2009. “How to build valid and credible simulation models”. In Simulation Conference (WSC), Proceedings of the 2009 Winter, 24–33. IEEE.

[13] Law A M. How to build valid and credible simulation models[C]//Proceedings of the 40th Conference on Winter Simulation. Winter Simulation Conference, 2008: 39-47.

[14] Gajski, D.D., S.Abdi, A. Gerstlauer, and G. Schirner. 2009. Embedded SystemDesign: Modeling, Synthesis and Verification. Springer Publishing Company, Incorporated.

[15] Harbola, A., D. Negi, and D. Harbola. 2012. “Infinite Automata And Formal Verification”. International Journal 2 (3).

[16] Gore R, Diallo S. The need for usable formal methods in verification and validation[C]//Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World. IEEE Press, 2013: 1257-1268.

[17] Wang, Y., The subversion server at the Technical University of Eindhoven, 2016. https://svn.win.tue.nl/repos/prom/Packages/TS2DEVS.

[18] Traoré M.K., SimStudio: a next generation modeling and simulation framework, Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008.

[19] Zeigler B.P., Hammonds P.E., *Modeling & simulation-based data engineering: introducing pragmatics into ontologies for net-centric information exchange*, Academic Press, 2007.

[20] Wang, Y., Zacharewicz, G., Chen D. and Traoré M.K., A proposal of using DEVS model for process mining, Proceedings of the European Modeling and Simulation Symposium, Bergeggi, pp. 403-409, 2015.