An underpinning theory and approach to applicability testing of constructive computational mechanisms

Li, Yongzhe; Horváth, Imre; Rusák, Zoltán

DOI
10.1007/s00163-022-00385-0

Publication date
2022

Document Version
Final published version

Published in
Research in Engineering Design

Citation (APA)
Li, Y., Horváth, I., & Rusák, Z. (2022). An underpinning theory and approach to applicability testing of constructive computational mechanisms. *Research in Engineering Design*, *33*(2), 213-230. https://doi.org/10.1007/s00163-022-00385-0

Research in Engineering Design (2022) 33:213–230
https://doi.org/10.1007/s00163-022-00385-0

ORIGINAL PAPER

An underpinning theory and approach to applicability testing of constructive computational mechanisms

Yongzhe Li[1] · **Imre Horváth**[2] · **Zoltán Rusák**[2]

Received: 8 January 2021 / Revised: 14 February 2022 / Accepted: 18 February 2022 / Published online: 26 March 2022

Abstract

Applicability testing of constructive computational mechanisms (CCMs) is a new challenge for both the academia and the industry. The overwhelming majority of the existing validation approaches focuses on the internal validity of CCMs (e.g. consistency, bias), while there is a shortage of efficient approaches for assessing the external validity (e.g. applicability, reusability). The objective of this paper is to clarify the concepts and criteria, and to develop an approach for a systematic evaluation of the applicability of a given CCM to cases that were not considered at design time. The approach is adapted from the validation square approach (VSA). The adapted methodology (A-VSA) makes it possible to evaluate CCMs from (a) theoretical structural, (b) empirical structural, (c) theoretical performance, and (d) empirical performance dimensions. Altogether eight indicators are introduced that support the evaluation process. The effectiveness of the A-VSA was confirmed through a case study, in which a specific CCM is considered and the strategy of the A-VSA was operationalized with three completely different application cases. As evidenced by the results, the proposed A-VSA establishes a tight coupling among the enablers embraced by a CCM and the aspects of theoretical and empirical validation, which approves the approach to be an efficient tool for defining the range and/or the extent of applicability. The advantage of the A-VSA is that it offers a way to transfer qualitative applicability evaluation into quantitative applicability assessment, which allows the use of both subjective statements and mathematical modeling in applicability testing. The results of the assessment can guide the adaptation work of a CCM when applied to an out-of-domain application.

Keywords Constructive computational mechanism · Systematic applicability validation · Structural appropriateness indicators · Practical appropriateness measures · Theoretical utility targets

Abbreviations

ART	Allowed response time
FEGS	Fire evacuation guiding system
A-VSA	Adapted validation square approach
I-CPSs	Informing cyber-physical systems
CAS	Care-taking assistant system
LTM/LTS	Practical latencies of information sensing/messaging
CCM	Constructive computational mechanism
OAI	Overall applicability indicator
DCIP-MM	Dynamic context information processing multi-mechanism
TMS	Traffic management system
ESI	Empirical suitability indicator
TPTs	Theoretical performance targets
EEI	Empirical efficiency indicator
TRI	Theoretical relevance indicator
FCS	Football-play coaching system
VSA	Validation square approach

1 Introduction

1.1 Introducing the addressed research issue

Constructive computational mechanisms (CCMs) are purposefully designed, implemented, and arranged algorithmic structures to support various knowledge-intensive activities (Parkes 2008). As main forms of knowledge inferring/reasoning, both non-ampliative and ampliative mechanisms have been proposed over the last decades.

✉ Yongzhe Li
yongzheli@seu.edu.cn

[1] School of Mechanical Engineering, Southeast University, Nanjing, People's Republic of China

[2] Faculty of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands

Application-specific reasoning mechanisms, which attempt to achieve a balance between application neutrality/specificity and performance dependability/efficiency, have also been considered (Horváth 2020). Currently, there is a paradoxical situation concerning the development of CCMs. The input information needed for development is typically obtained by considering only one or a limited number of potential application cases, while the broadest possible range of applications is expected when the mechanism has been developed. The issue is that a CCM tailored to a specific application may not match procedurally or may show performance deficiencies in nonconforming applications, or may completely fail in borderline applications (Debbabi et al. 2010).

For instance, most reinforcement learning-based computational mechanisms for robot path planning are applicable in static working environments but only efficient to cases where adequate training efforts are performed (Zhao et al. 2020). Then incongruity between the development and application of CCMs negatively influences software and algorithms reusability and the efficiency/economy of development. Accordingly, investigation of the applicability of CCMs is a primary and central objective of validation efforts in system engineering, and an important issue for both academia and industry (Brazdil et al. 1994; Barambones et al. 2021).

1.2 Elaboration on the concept of applicability validation

The term 'validation' is frequently used intuitively or ambiguously in the literature. It seems that theoretical and pragmatic interpretations coexist (Pardo 2016). The term is often used to depict activities that belong to the scope of 'verification' or 'consolidation' of scientific theories and/or other research results. Furthermore, authors often replace the term 'validation' with synonyms such as certification, attestation, authentication, or confirmation. Therefore, disambiguation of the notion and explanation on how it can be used in the confirmatory phase of research is necessary and useful. In the tradition of scientific inquiry, validation is a multi-faceted activity focusing on the confirmation of knowledge. As discussed by Barlas and Carpenter (1990), while verification refers to internal consistency, validation refers to the appropriateness of knowledge claims. In the process of establishing scientific theories, validation is done to test and prove appropriateness and utility for a purpose in some (application) context (Donald 1995).

The pragmatic interpretation of validation is about checking the fulfillment of some expected functionality and fit for purpose (Pape et al. 2013). In simple words, the main question of validation is whether a new research result (body of knowledge, theory, framework, methodology, etc.) does what it is supposed to do? The reason why it may not happen can be biases and errors in the conduct of research or lack of adequacy. Thus, validation should focus on both the critical factors (possible sources and forms of biases) and the appropriateness of the results (findings) of research in particular contexts (e.g. in various real-life situations). Consequently, internal validation and external validation are distinguished (Smolka et al. 2009). While internal validation aims at exploring and evaluating biases, external validation checks issues related to generalizability and reusability.

In the literature, applicability validation is not among the most frequently addressed aspects of validation. For this reason, its methodological support is underdeveloped. The shortage of efficient testing approaches makes applicability validation challenging. As evidenced in the literature, the development of applicability validation methodologies typically considers a limited number of application cases only, whereas the developed computational mechanism is supposed to cover the broadest possible range of application cases. This paper was intended to resolve this paradoxical situation.

1.3 Overview of the latest developments in systematic and rigorous validation

Validation is defined as the substantiation that a computerized model possesses a satisfactory range of accuracy consistent with the intended application of the model within its domain of applicability (Schlesinger 1979). Based on our extensive literature study, five subject areas for validation have been identified, namely, validation of (i) data, information, and knowledge, (ii) concepts, theories, models, (iii) objects, structures, and systems, (iv) actions, processes and services, and (v) methods, methodologies and tools. This categorization served as a kind of reasoning model for the conducted work. However, due to the space limitation, only the contributions to the subject area of validation of methods, methodologies, and tools were summarized, where CCMs are related.

Method validation is one of the universally recognized challenges of any rigorous research (Frey et al. 2006). For this reason, the number of techniques, protocols, and guidelines for research method validation is large. The general objective is to demonstrate whether a method is fit and effective for a particular constructive or analytical purpose. Teegavarapu (2009) argued that case study-based development of design method enables in-depth analysis in real-life contexts. Kroll and Weisbrod (2020) proposed that a plurality of measures can be jointly used to assess the applicability and effectiveness of design methods. In engineering contexts, methods have been proposed for both empirical and virtual validation. For instance, Mejía-Gutiérrez and Carvajal-Arango (2017) discussed (i) abstract prototypes, (ii) virtual prototypes, (iii) functional prototypes, and (iv) physical prototypes as the four types most frequently used in the design validation stage of product development.

CCMs are sophisticated enablers of computational reasoning in various application cases. As such, they need systematic approaches to have their appropriateness and performance tested in the concerned applications. Eze et al. (2011) discussed various challenges associated with validation of self-managing and autonomic computational mechanisms and investigated the relations of various validation approaches (such as system unit testing, real-world system testing, pervasive supervision, system model checking, and system self-testing) and execution modes (generic, design time, run-time, integrated, and autonomous). Ahmad et al. (2015) analyzed validation techniques for safety–critical software for the purpose of (i) functional failure analysis, (ii) hazard and operability studies, (iii) failure modes and effects analysis, and (iv) fault tree analysis. They compared these techniques according to their (i) efficiency, (ii) reliability, (iii) dependability, (iv) testability, and (v) usability. Brings et al. (2016) dealt with the issue of supporting early validation of CPS specifications by model-based prototype development.

Gonzalez and Barr (2000) exposed the differences to be considered in the verification and validation of intelligent systems. Feth et al. (2015) focused on the validation of open and heterogeneous systems, such as CPSs in the automotive domain, and proposed a simulation-based framework, which integrates AUTOSAR applications. The virtual validation concerned the functional behavior and the performance of the software. Guarro et al. (2016) proposed a comprehensive, multi-level framework for validation of model-based control and adaptive control systems, which combines logic dynamic model constructs and the associated analysis processes to demonstrate compliance with the related aviation-system certification standards. Olewnik and Lewis (2005) elaborated on the validation of design decision methodologies. According to them, the complexity of prescriptive models makes their validation a difficult task.

Complex systems usually comprise a large number of interacting modules, which require both module-level validation and functionality-level validation (Fu et al. 2015). As claimed by Christophe et al. (2010) and Reich (2017), the current performance characteristics of complex systems and mechanisms are usually evaluated depending on the requirements that are made in their initial design stages. Zheng et al. (2016) proposed a multi-disciplinary interface model to ensure consistency and traceability between system-level and functionality-level functionalities for verification and validation of mechatronic systems. Dauby and Dagli (2011) proposed a methodology for the assessment of the system of systems using general system attributes or overall strength/weakness metrics.

According to the survey work, there are both experimental and analytical (logical) approaches to validate computational methods. The approaches are logically rigorous, internally consistent, and mathematically correct. However, the literature offers only limited insights with regards to the applicability validation of CCMs. On the one hand, the criteria for applicability validation of CCMs are still unclear. On the other hand, the existing validation work is largely influenced by the purpose of their applications, which addresses internal validity. As a typical aspect of external validity, applicability validation focuses on the external appropriateness of CCMs. These two challenges jointly stimulate this scientific research.

1.4 The focus and contents of this paper

This paper proposes basic concepts and criteria and develops an approach to applicability validation of constructive computational mechanisms, which are validated in a dedicated case study. The next section introduces the essence, reasoning model, and procedural principles of the developed approach. Section 3 presents an already developed computational mechanism and shows how its applicability is evaluated using the proposed approach. Section 4 reflects on the results of the work. Conclusions are given in Sect. 5.

2 Towards applicability testing of constructive computational mechanisms

The validation square approach (VSA) proposed by Pedersen et al. (2000) is a theoretically well-established and methodologically reasonably transparent validation approach that concerns the external validity of design methods. This approach combines qualitative and quantitative assessments, which lends itself to the realization of an effective, content, and context-independent applicability validation approach. It has been studied from both epistemological and procedural perspectives and adapted to the needs of applicability testing of design methods through multiple cases, such as the work reported by Du Bois and Horváth (2013). Based on these foundational considerations, we adapted the initial reasoning model of VSA and made it better fitting for applicability validation of CCMs. This approach will be referred to as A-VSA in the rest of the paper.

2.1 The validation square approach and its features

The primary interest of the VSA was “How does one verify and validate a design method?” Being aware of the facts that (i) knowledge is associated with heuristics and non-precise representations in the process of design method synthesis, and (ii) that validation of new knowledge becomes a process of building confidence in its usefulness concerning a purpose, the author epistemologically associated verification with internal rational consistency (logical properness), while validation with confirmation of knowledge claims (utility of solution). Later on, Seepersad et al. (2006) asserted that a formal logic-based, rigorous and quantifiable validation can be applied to a design method to evaluate its internal consistency, but an approach like this would fail to validate its external relevance (i.e., its usefulness). Consequently, they argued that a design method should be considered to be validated if it could be proven useful for the intended specific purpose.

The above considerations have been extended with other ones concerning the nature of validation and the objective of validation. With regards to nature, theoretical and empirical dimensions have been differentiated. Theoretical validity deals with the validity of a design method for a general problem. Empirical validity concerns the validity of a design method for specifically chosen examples. With regards to objective, structural validity and performance validity have been considered. Structural validity involves qualitative testing of the logical structure of a design method for correct results. Performance validity involves quantitative testing of the capability of a design method to produce useful results. These four dimensions give a framework of VSA (Fig. 1). Each half of the square is further divided into a domain-specific and a domain-independent quadrants, associated with the validity of the method for the domain-specific examples investigated in the research and for broader domains of application, respectively.

The quadrants of the validation square define the aspects of validations, and the square combines them in a particular semantic and procedural arrangement, as follows:

(i) Theoretical structural validity deals with the internal consistency of a design method and checks the logical soundness of the constructs both individually and integrated;

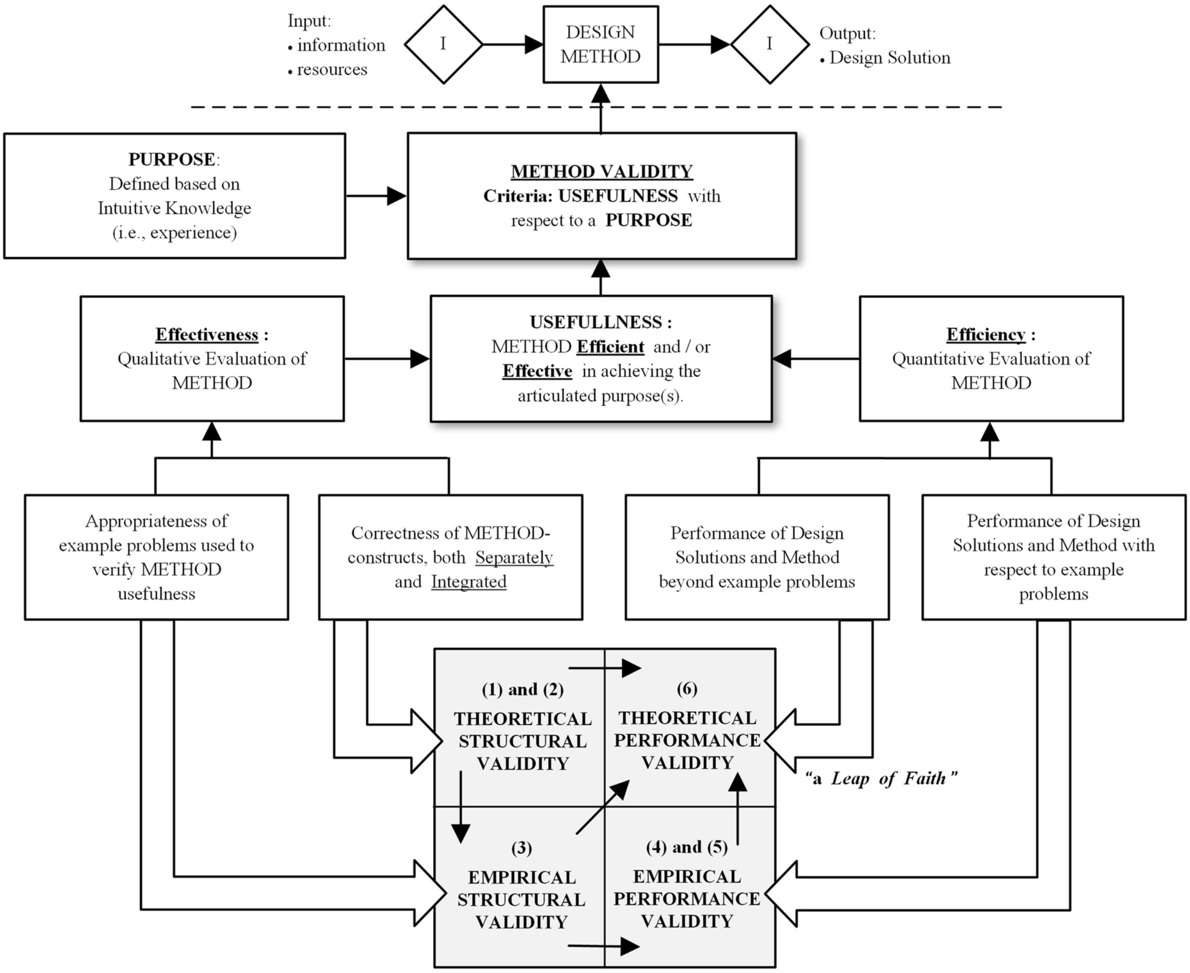


Fig. 1 The reasoning model underpinning the validation square approach (Pedersen et al 2000)

(ii) Empirical structural validity deals with the appropriateness of the design method to chosen example problem(s) with the intention of having correct results;
(iii) Empirical performance validity concerns the ability of a design method to produce useful results for the chosen example problem(s); and.
(iv) Theoretical performance validity indicates the capability to produce useful results beyond the chosen example problem(s).

Procedurally, the VSA can be realized by executing the steps (i) ~ (iv) in the order of mention. As the authors emphasized, the built confidence in the general usefulness of a method requires a 'leap of faith'. Epistemologically, the four quadrants make it possible to argue about domain-specific performance validity (e.g. if the method is useful in a specific sense) and about domain-independent performance validity (e.g. if the method is found useful in a more general sense). However, a limitation of the VSA is that it only offers a philosophical view on validating design methods, while it does not offer specific methods/tools of validation (Teegavarapu 2009). VSA has some other limitations such as (i) validity of a body of knowledge, that conveys the know-how of a design method, is determined by many more measures than just its usefulness, (ii) validation should explore and reduce biases and errors (increase credibility by internal validation), and (iii) validation should test potentials and implications in varying contexts (transferability delivered by external validation). Nevertheless, usefulness is difficult to disprove as a pragmatic objective of validation of methods, methodologies, and tools.

2.2 The reasoning model for applicability testing of CCMs

Our research focused on the exploration of the potential of VSA as well as putting it into a different logical and procedural framework. Accordingly, the fundamental assumption was that, if the internal consistency of a CCM is guaranteed, applicability validation is the process of building confidence in its appropriateness and usefulness concerning an application purpose. As a criterion of applicability, appropriateness indicates if the concerned CCM can produce utility, and usefulness indicates that it can enable the embedding system to work correctly. Appropriateness and usefulness are considered in the various quadrants of the validation square in the following way: (i) from a theoretical structure perspective, appropriateness is captured by the indicators of relevance, (ii) from an empirical structure perspective, appropriateness is captured by the indicators of suitability, (iii) from an empirical performance perspective, usefulness is captured by the indicators of efficiency, and (iv) from a theoretical performance perspective, usefulness is captured by the indicators of sufficiency. For each quadrant, two questions are obtained from the intersected aspects of validation. Figure 2 shows the reasoning model for using the approach in the applicability validation of CCMs.

A CCM can be effectively applied in a target application, if the structural construct specified by the CCM from a procedural perspective (the way the constructs are put together in an executable process) and the relationships among these constructs (the logical and functional dependencies between them) match (appropriate for) what are requested by the target application. This can be captured by a measure called 'relevance'. As a concrete representation of the theoretical structure proposed by a CCM, functionalities and workflow charts can be used to capture all computational transformations and their interrelationships (interoperation). If the functionalities and workflow chart required by the new (target) application are developed, then both can be compared with the CCM, and the congruencies and the differences can be identified and evaluated. These aspects can be used to decide on the range of applications for which the CCM is appropriate without or with limitations.

In the case of CCMs, the sufficiency condition of structural validity is that the mechanisms and algorithms proposed by the validation methodology, as well as the data, information, and knowledge constructs and structures processed by the mechanisms and algorithms, are appropriate for the particular application. This can be referred to as empirical structural validity and captured by a measure called 'suitability'. This seems to be a significant deviation from the initial interpretation of the VSA, in which the appropriateness of the example problems used to verify the performance of a concerned design method was assessed and it was deemed to be one of the actions related to theoretical structural validity. To check the practical structure, appropriateness of algorithms, and data constructs, respectively can be used. This way, comparison and evaluation of what is proposed by the CCM and what is required by the target application can inform about the measure of empirical structural validity. In other words, it informs about whether the CCM can compute what is needed in the target application case and whether the data/knowledge needed by the CCM can be provided and the output is an adequate input to other mechanisms and algorithms in the target application case.

The practical performance of CCMs depends on multiple factors such as (i) the efficiency of the execution process, (ii) the choice of methods, (iii) the selection of means, and (iv) the criteria of utility. This implies the need for reconsideration of the performance validation strategy of the original VSA. Accordingly, evaluation of the usefulness of a CCM was proposed to be a quantitative process centered on 'efficiency'. Empirical performance validity is related to the modeling and computing

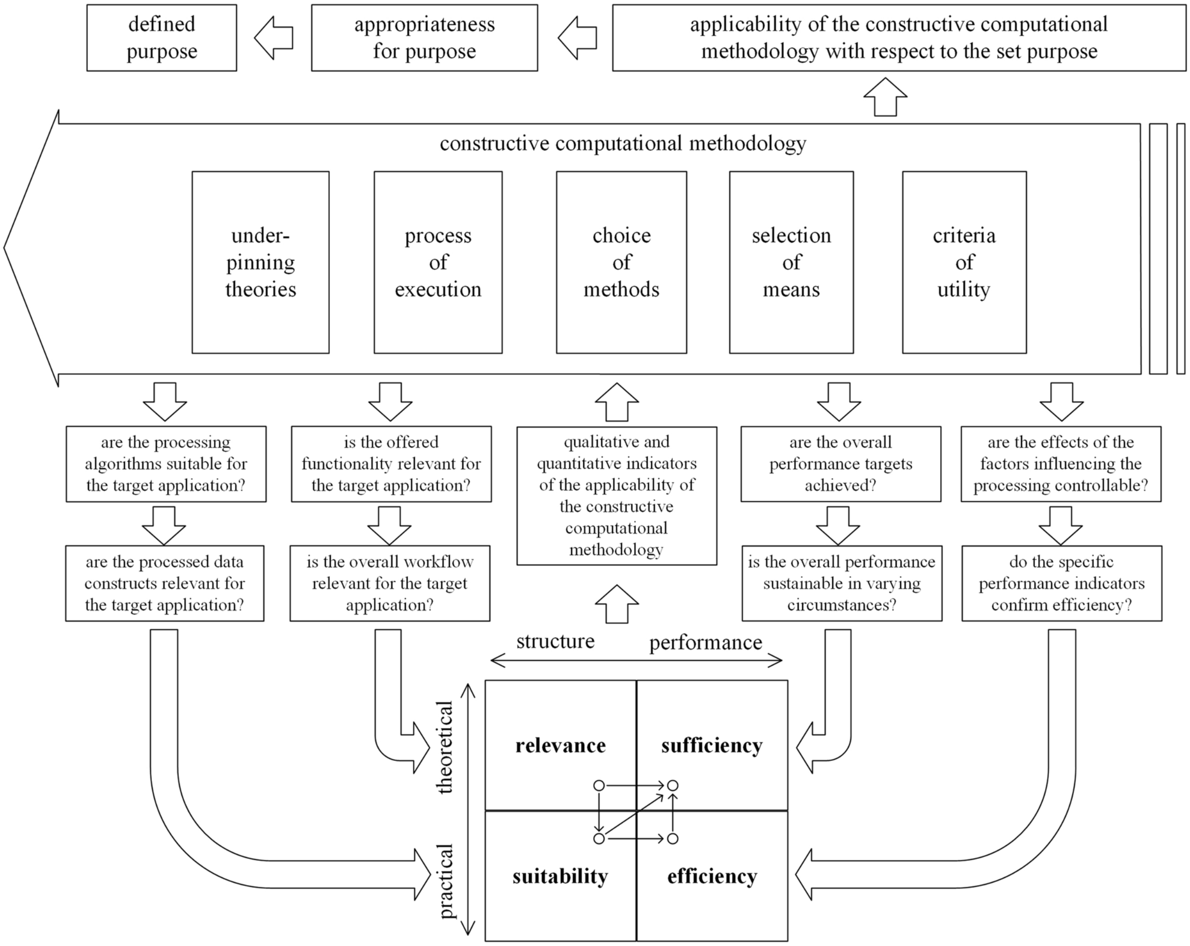


Fig. 2 The adapted reasoning model used to underpin the validation square approach in applicability validation of constructive computational mechanisms

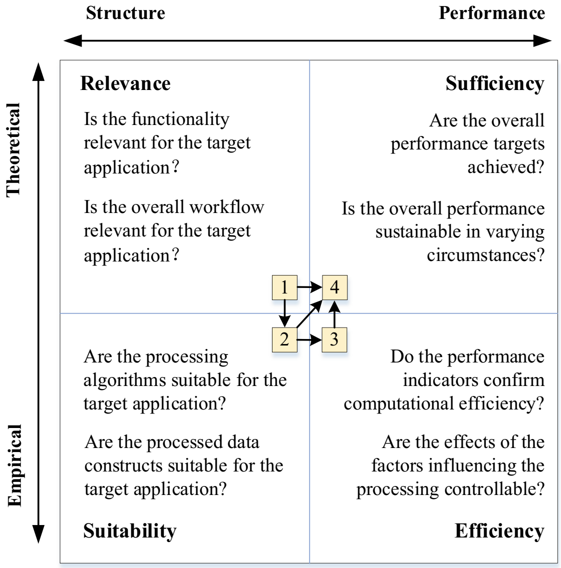
performance of involved computational mechanisms and algorithms, and to the reliability, completeness, and computability of the data, information, and knowledge structures/constructs. Measures such as (i) the amount of input data, (ii) the time of data preparation, (iii) the simplicity of models, (iv) the input efforts, (v) the computation time, (vi) the amount of output data, and (vii) the sensitivity of the outcome to errors have been used as concrete quantitative indicators of efficiency. In addition, various non-numerical indicators were also considered, for instance, (i) the ease of use, (ii) the human cognitive effort, (iii) the trust-building by individuals, (iv) the chance of making errors, (v) the transparence of inputs/outputs, and (vi) the awareness of the work of computational mechanisms and algorithms. Typically, these need a qualitative description or characterization. It entails that empirical performance validation of the applicability of CCMs can be done not only quantitatively, but also qualitatively, and/or in combination of the two.

Theoretical performance validation of a method is driven by its defined purpose. The main question is if a CCM can achieve certain defined performance targets. The targets are objective expectations about the overall performance of ‘sufficiency’ in the target application case. The theoretical performance targets can be (i) to reduce costs, (ii) to save time, (iii) to reduce efforts, (iv) to eliminate errors, (v) to resolve bottlenecks, (vi) to complement knowledge, (vii) to facilitate collaboration, (viii) to stimulate innovation, and (ix) to improve quality. In addition, the mentioned targets can also be used in circumstances other than the one that provided the input requirements and the information for conceptualization and development of the particular validation methodology. Even if the purpose of applying a CCM in a given application may be largely different, there is a possibility to formulate general performance targets that are more or less application independent. This enhances the external validation nature of the original VSA.

2.3 Proposing an adapted-VSA to applicability testing of CCMs

Based on the proposed assumption and criteria, the flow of knowledge elicitation for application validation, including the four indicators (relevance, suitability, efficiency, and sufficiency) and the questions formulating the fundamental criteria are shown in Fig. 3. In the process of evaluation of applicability, the A-VSA considers: (i) the relevance of the structural constituents of the CCM, (ii) the suitability of the empirical procedural/information flow and data constructs, (iii) the efficiency of the empirical performance enablers as indicators and measures, and (iv) the sufficiency of the theoretical performance targets. The existence of a tight coupling among the four quadrants can be observed. Furthermore, the aspects of applicability validation are interrelated through certain semantical and functional dependencies.

As a starting point, the theoretical structural validity is supposed to confirm the internal appropriateness of the CCM to the target application case. This forms the basis of the empirical structural validity (the arrow from quadrant 1 to quadrant 2). What this interrelationship means is that there must be a match between the theoretical structure arranging the constituents and the concrete computational components of the used CCM. If this match cannot be shown, then the concerned CCM cannot be applied to the particular application case. The criterion of structural appropriateness can be met if the algorithms and the logical and functional dependencies between them, as well as the data, information, and knowledge constructs/structures processed by the algorithms, are appropriate for the particular application. However, the theoretical structural validity is only a necessary condition, but not a sufficient condition of applicability. The theoretical and the practical performance should also be appropriate.



Relevance

Is the functionality relevant for the target application?

Is the overall workflow relevant for the target application?

Sufficiency

Are the overall performance targets achieved?

Is the overall performance sustainable in varying circumstances?

Suitability

Are the processing algorithms suitable for the target application?

Are the processed data constructs suitable for the target application?

Efficiency

Do the performance indicators confirm computational efficiency?

Are the effects of the factors influencing the processing controllable?

Fig. 3 The strategy of applicability validation using the adapted validation square approach

The evaluation of the empirical performance can happen only if the empirical structural validity is confirmed (the arrow from quadrant 2 to quadrant 3). It has to be mentioned that the usefulness of a design method was initially linked to the empirical performance validity. We assumed that, in addition to assessing whether the specific quantitative and qualitative performance indicators confirm efficiency, it is also important to investigate if these can be maintained under the effects of the factors influencing the controllability of the computations and data processing. For instance, the influential factors can be (i) the availability of cloud resources, (ii) the incidental malfunctioning of some algorithms, (iii) the logical fallacies in reasoning, (iv) the temporal incompleteness or incorrectness of data, (v) the sensitivity of data/computation to deviations, and so forth. For this reason, empirical performance validation should have a strategy for identifying the most critical influencing factors and procedures, and for their ranking based on the evaluation of their impacts on the values of the empirical performance indicators. Investigation of the factors influencing the outcome of applying the methodology is useful from the point of view of building confidence in the applicability of a computational mechanism.

Likewise, the efficiency of empirical performance is only a necessary condition for usefulness. The sufficient condition is that all theoretical performance targets should be achieved (the arrows from quadrant 1, 2, and 3 to quadrant 4). The theoretical performance targets are objective expectations concerning the overall performance of the computational mechanism in various applications. On the one hand, the indicators are chosen to provide a reliable characterization of applicability in real-life application cases. On the other hand, they are also supposed to hint at necessary or possible adaptations towards high-level appropriateness and usefulness. The indicators can be chosen to provide either qualitative (interpretative) or quantitative (nominal or statistical) characterization, or both. This is a difference in comparison with the original VSA, in which the theoretical and empirical structural 'validities' were supposed to be evaluated only qualitatively.

3 A case study: validation of the A-VSA

In this section, we provide a brief overview concerning how the proposed criteria and the A-VSA can be applied to a specific case. This section first introduces an already developed CCM with a reference application that the CCM can perfectly handle. Then, three completely different application cases are presented that the CCM has the potential to be applied for. Details about the applicability evaluation procedures are also discussed.

3.1 Introducing a developed CCM

To support various knowledge-intensive decision making and action planning activities of informing cyber-physical systems (I-CPSs), a dynamic context information processing multi-mechanism (DCIP-MM) has been developed. The overall view of the DCIP-MM is shown in Fig. 4. The DCIP-MM includes four modules (sub-mechanisms), which are dedicated to (i) dynamic context information representation, (ii) semantic context information inferring, (iii) context-based real-life action-plan generation, and (iv) context- and receiver-sensitive message generation and sending. A brief introduction of the mechanism is as follows: After control commands given by the I-CPSs are confirmed, the information representation module (M1) acquires time-dependent descriptive context data about entities and relations. The data were used to construct CIR-cubes, which is a context representation scheme in the DCIP-MM. In the knowledge inferring module (M2), situations are inferred based on the CIR-cube and their impact indicators on involved entities are calculated. Then, the inferred situations and impact indicators together form a basis for treating the situations towards generating personalized solutions and action plans for context management in the action-plan deriving module (M3). Based on the determined personal action plan and the inferred semantic knowledge, the message constructing module (M4) generates personalized messages and the order of message sending. Further information about the theoretical fundamentals, conceptualization, and implementation of the DCIP-MM can be found in (Li 2019).

Concerning the development of the DCIP-MM, the intention was to create an application-independent computational platform that supports certain reasoning and messaging tasks expected from I-CPSs, and that (i) is adaptable to various application tasks, (ii) can provide high-level flexibility in operation, and (iii) broadens the range of the addressable computational problems. The conceptualization and implementation of the DCIP-MM needed requirements from a specific application. For this reason, a fire evacuation guiding application case was considered as a reference. Many algorithms of the DCIP-MM are completely or largely application-independent, but some algorithms (e.g. generate instructions about fire situation) and parameters (e.g. threshold) feature application dependence. The fact that the DCIP-MM is intended to be used in several different applications of I-CPSs raised the issue of applicability validation and necessitated the development of a dedicated validation methodology.

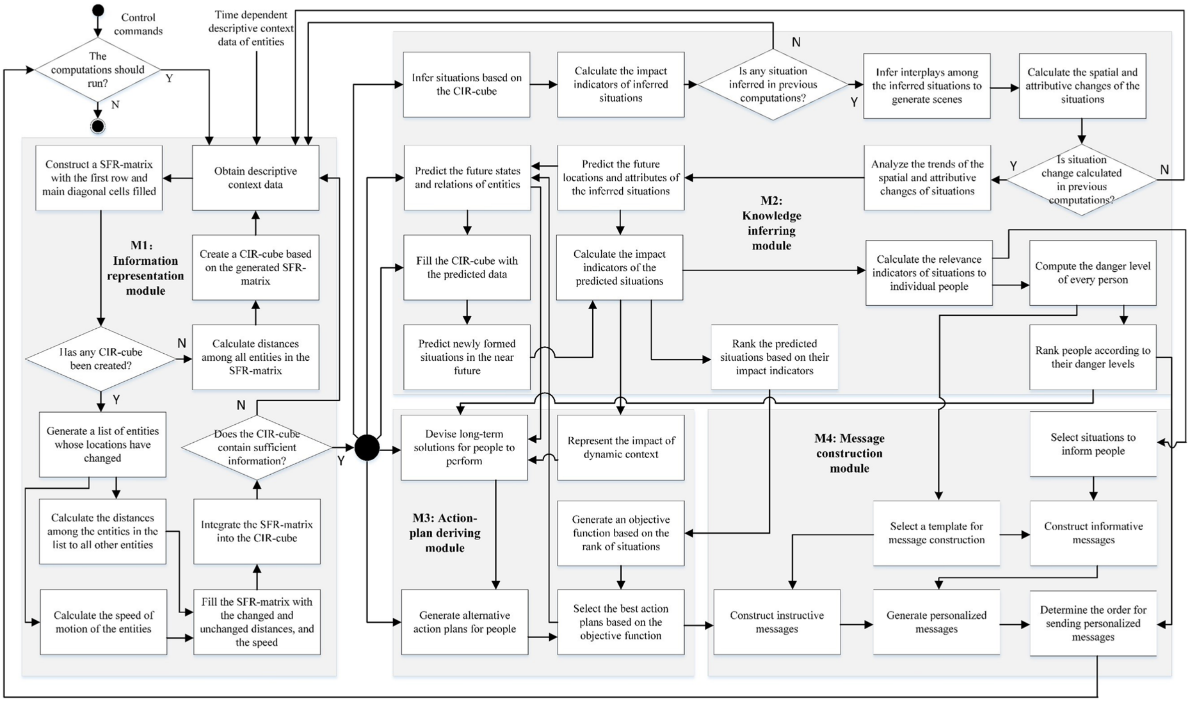


Fig. 4 High-level overview of the specific functionalities included in a DCIP-MM

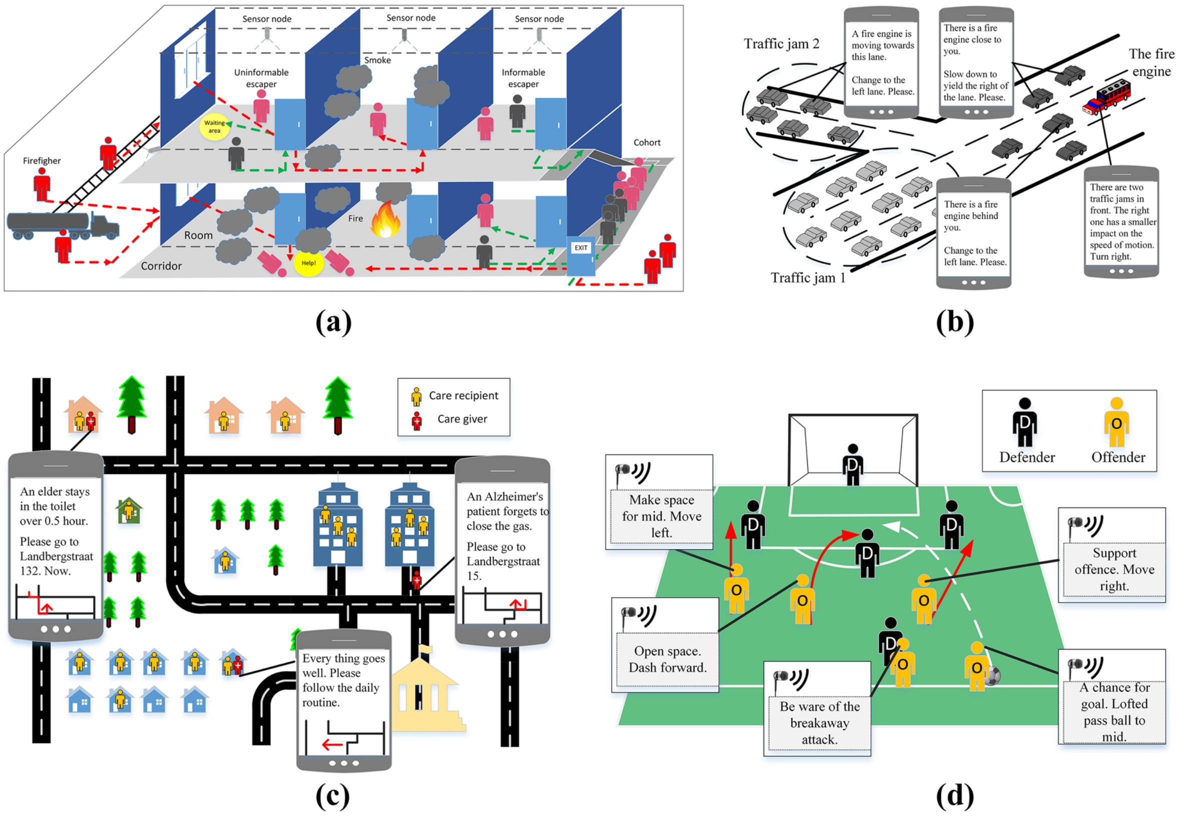
3.2 The reference application case

The proposed DCIP-MM has been developed with an indoor fire evacuation guiding system (FEGS) in mind. It was used both as a computational scenario and as a source of application data/knowledge. The FEGS aims at providing personalized escape guidance for people in different situations (Horváth et al. 2017; Li et al. 2019). A symbolic image of the FEGS is shown in Fig. 5(a), and its main tasks were: (i) understanding the phenomenon of indoor fire evacuation, (ii) constructing representation schemes for spatiotemporal context data, (iii) deriving semantic information and knowledge based on dynamic context information management, (iv) developing situation-dependent and personalized escape routes for individuals in a quasi-real-time manner, (v) sending informative and instructive messages to all individuals having communication possibility, and (iv) adapting to the obedience of message receivers.

The DCIP-MM was developed to generate operational strategy and to synthesize personalized action plans for all involved individuals, no matter if they were directly (through their smartphone) or indirectly (through the involvement of other individuals) notified about these plans. The DCIP-MM determines the best route for the concerned person to escape, updates the individual escape action options, sends information about this to each person, monitors the obedience of informed individuals, and adapts to the dynamic contextual changes. The DCIP-MM was used not only in the development of the dynamic context computation, action-plan generation, and message construction and distribution mechanisms but also in their testing. A high-fidelity simulation of (i) the propagation of the fire and (ii) the behaviors of human, artefactual, and natural entities was applied to correctly reproduce the presumed real-life cases. The obtained experimental results proved the efficiency of the computational mechanisms and the interoperating algorithms. They also confirmed that the proposed DCIP-MM can provide both descriptive and predictive knowledge about emergencies as well as about the implications of the interplaying situations on the entities in quasi-real-time.

3.3 Three target application cases

Called target applications, three possible but not trivial I-CPS applications have been considered. Each of them needs dynamic context information processing, situated



A fire engine is moving towards this lane.

Change to the left lane. Please.

There is a fire engine close to you.

Slow down to yield the right of the lane. Please.

There is a fire engine behind you.

Change to the left lane. Please.

There are two traffic jams in front. The right one has a smaller impact on the speed of motion. Turn right.

An elder stays in the toilet over 0.5 hour.

Please go to Landbergstraat 132. Now.

An Alzheimer's patient forgets to close the gas.

Please go to Landbergstraat 15.

Every thing goes well. Please follow the daily routine.

Make space for mid. Move left.

Open space. Dash forward.

Be ware of the breakaway attack.

Support offence. Move right.

A chance for goal. Lofted pass ball to mid.

(a) (b) (c) (d)

Fig. 5 The reference application and target application cases considered for applicability testing: **a** indoor fire evacuation guiding system (FEGS), **b** traffic management system (TMS), **c** care-taking assistance system (CAS), and **d** football-play coaching system (FCS) reasoning, and personalized message communication, which lend themselves to novel applications of a smart I-CPS.

Target application 1: The traffic management systems (TMS) are deemed to be necessary and useful due to the increase in the number of vehicles on city roads. As shown in Fig. 5(b), a specific traffic management scenario was considered as follows. A fire engine is moving on the road for a fire distinguishing mission. There are two traffic jams accidentally formed on the highway, which may influence the movement of the fire engine. The objective of the TMS is to determine the optimal routes and actions for the related drivers to enable the fire engine to arrive at the target location as fast as possible. Processing dynamic context information is necessary since the possible rapid changes of the state and situation of the individual vehicles (e.g. their motional attributes, the distance between them, etc.) have an impact on the decision-making process of the drivers. To manage the dynamic context of traffic situations, the TMS should: (i) continuously monitor the emerging situations on the road, (ii) extract and manage the spatial, attributive, and temporal data of the vehicles as well as the data about their spatiotemporal relations, (ii) infer high-level context knowledge and meta-knowledge about the situations, their impacts on relevant entities, (iii) prioritize the entities and situations and apply the prioritization in the decision-making process, (iv) derive action plans for the drivers that are involved in a particular situation taking into account the overall system objectives, and (v) construct personalized messages to the drivers according to the action plan they should follow (including all pieces of information they should be aware of).

Target application 2: A home care-taking assistant system (CAS) provides caretaker services for care recipients (e.g. elderlies, patients, etc.) in a home environment. A specific scenario for the application is shown in Fig. 5(c). Three caregivers provide daily care for a group of people. As their routine tasks, each of the caregivers has to look after several people in a day and their moves should be done with a minimal totaled motion path. However, in the case of emergent situations and circumstances, they have to act according to the incurred level of danger. If more than one dangerous situation happens at the same time, then they have to make decisions about an action plan. To support the caregivers' work, the CAS should be involved in (i) monitoring the change of conditions, (ii) managing the schedule, (iii) reasoning with the servicing capacity of the caregivers, and (iv) providing action plans and information in real-time. To achieve these, the CAS should process dynamic context information concerning the caregivers and the care recipients. Accordingly, the CAS is supposed (i) to aggregate information about the personal context of care recipients (e.g. activities) as well as about the caregiver (e.g. location), (ii) to infer knowledge about any dangerous situation happening in or around the home of the care recipients, (iii) to evaluate the relevance of the situations to caregivers, (iv) to derive action plans for the caregivers to follow (including the order of caretaking and the needed actions), and (v) construct personalized messages (including informative messages and instructive messages) for taking care of the care recipients.

Target application 3: A real-time football-play coaching system (FCS) is supposed to real-time contact players of a team simultaneously to improve group performance. Figure 5.d shows a hypothetical scenario. The offenders shown are given personalized messages about the best momentarily strategy and collective actions to score a goal. The information provisioning can happen through wearables and/or portable equipment (e.g. mini wireless headset). Since the offenders may play different roles (e.g. scorer, supporters, and passer), the information or instruction can be selectively formulated according to the dynamically changing contexts. In this scenario, the dynamic context of the football player includes (i) the changes of the attributes of the other players (e.g. location, speed of running, and physical agility), and (ii) the changes of the spatial relations (e.g. distances and orientations) among a concerned player, the other players, and the ball. To advise the players with proper information, the FCS should (i) handle time-dependent information concerning all players and the ball on the playground, (ii) infer and predict situations (e.g. an open corridor, a blockage, a large space behind the defenders), (iii) forecast the implications of the situations on players, (iv) derive action plans that can be used at suggesting the players what to do, (v) construct personalized messages for the players (e.g. quick counterattack, offense, defense, etc.), and (vi) adapt to the contextual changes (e.g. if players obey the instructions or not).

To support evaluation of the empirical validity, the enabling technologies applied by the chosen target applications may involve different constraints, e.g. communication and computation hardware. The realization of expected performance depends on the empirical influencing factors. Therefore, the characteristics and influencing factors of the chosen target applications were focused, to create a factual basis for the applicability analysis and validation. Three major characteristics for comparative applicability analysis were selected, namely (i) the number of entities handled at a given point in time or in a specific time interval (computational time-increment), (ii) the response time required by the 'happenings' in the application case, and (iii) the sudden change of entity number involved between two successive intervals. The specific quantitative values associated with these characteristics are indicated in Table 1. Although there might be some exceptional situations, we assumed that these values could properly fit most cases in the application contexts. In addition, eight factors influencing information sensing and messaging were considered to represent the technical constraints in the chosen application cases, which are shown in Table 2. The technological and operational comparison showed that the empirical characteristics and the concerned aspects of influencing factors were rather varied in the selected target application cases. Nevertheless, each of these applications could be considered as a representative of a family of slightly different application cases sharing similar characteristics and situations.

3.4 Validating the applicability of the DCIP-MM to the target applications

3.4.1 Analysis of the theoretical structure relevance

The investigation of the theoretical procedural structure included the following activities: (i) specification of the exemplified procedural structures of the considered I-CPS applications, (ii) specification of the indicators of applicability, and (iii) comparing the proposed and the required procedural structures. It characterized the relevance of the functionalities of the DCIP-MM and the relevance of the processing workflow of the DCIP-MM to the traffic management, care-taking assistance, and coaching football playing applications. The results of the assessment of the relevance of the functionalities for the target applications are shown in Table 3. The relevance of the functionalities was qualitatively evaluated using three values: (i) very relevant, which meant that the functionality can fulfill all the requirements of the target application, (ii) partially relevant, which meant that the functionality can partially fulfill the requirements of the target application, and (iii) not relevant, which meant that the functionality cannot fulfill any requirement of the target application at all.

To evaluate the relevance of the workflow, functionality dependence was used to characterize the relationship of functionalities in a workflow. For a given functionality that was needed by an application case, its previously

Table 1 A comparison of the application-implied requirements of the applications

Characteristics	FEGS	TMS	CAS	FCS
The number of entities handled at a time	40–300	100–1000	20–100	23
Required response time	Less than 5 s	Less than 1 s	Less than 10 s	Less than 0.5 s
The sudden change of entity number	0–10	10–200	0	0

Table 2 Comparison of the technical factors that influence the empirical performance of the application cases

Influencing factors	Application cases			
	FEGS	TMS	CAS	FCS
Used sensing technology	UWB	GPS	GPS + Indoor camera	Camera
Typical positioning accuracy	Less than 0.5 m	Less than 4.9 m	Less than 4.9 m	Less than 0.1 m
Latency of information sensing	Less than 0.25 s	Less than 0.5 s	Less than 0.5 s	Less than 0.1 s
Power supply of terminals	hand-held	on-board	hand-held	in-ear
Technology for message sending	WLAN/WiFi	4G Mobile web	4G Mobile web	Bluetooth 5.0
Latency of message receiving	Less than 20 ms	Less than 100 ms	Less than 100 ms	Less than 3 ms
Length of personalized messages	Less than 30 words	Less than 30 words	Less than 80 words	Less than 10 words
Latency of message reading	0-∞ s	0.5 s	0-∞ s	0.5 s

Table 3 Assessment of the relevance of the functionalities for the target applications

No	TMS	CAS	FCS	No	TMS	CAS	FCS	No	TMS	CAS	FCS	No	TMS	CAS	FCS
F1.1	●	●	●	F2.1	●	●	●	F2.10	●	●	●	F3.5	●	×	●
F1.2	●	●	○	F2.2	●	●	●	F2.11	●	●	●	F4.1	●	×	●
F1.3	●	×	○	F2.3	●	×	●	F2.12	●	●	●	F4.2	●	●	×
F1.4	●	●	○	F2.4	●	×	●	F2.13	●	●	×	F4.3	●	●	●
F1.5	●	×	×	F2.5	●	×	●	F2.14	●	●	×	F4.4	●	●	●
F1.6	●	●	×	F2.6	●	●	●	F3.1	●	×	×	F4.5	●	●	●
F1.7	●	×	○	F2.7	●	●	●	F3.2	●	●	×	F4.6	●	×	×
F1.8	●	●	○	F2.8	●	●	○	F3.3	●	×	●				
F1.9	●	●	○	F2.9	●	●	●	F3.4	●	●	●				

●: Very relevant, ○: Partially relevant, ×: Not relevant performed functionalities defined the dependency. Accordingly, a comparison of the dependency of the functionalities for the concerned applications is shown in Table 4. The assessment considered three values: (i) same dependency, which means that the functionalities performed before a concerned one in the target application are the same as those specified in the DCIP-MM, (ii) different dependency, and (iii) the functionality is not included in the workflow. The dependency of the functionalities for traffic management was the same as in the case of the fire evacuation guiding. It meant that the workflow could be reused directly. For the CAS and the FCS, certain functionalities specified for the fire evacuation were not needed. This fact negatively influenced the reusability of the related workflow. For details about the assessment process, please refer to (Li 2019).

3.4.2 Analysis of the empirical structure suitability

The required functionality is realized by the algorithms integrated into the given CCM and the related data constructs. The analysis of the suitability of the empirical structure of the DCIP-MM was completed in two steps. Indicators were derived with regards to the suitability of the algorithms incorporated by the DCIP-MM as well as to the suitability of the data constructs to the application in traffic management, care-taking assistance, and football-play coaching applications. There are 8, 21, 10, and 6 algorithms in the

Table 4 A comparison of the dependency of the functionalities for the concerned applications

No	TMS	CAS	FCS	No	TMS	CAS	FCS	No	TMS	CAS	FCS	No	TMS	CAS	FCS
F1.1	●	○	●	F2.1	●	●	●	F2.10	●	●	●	F3.5	●	×	○
F1.2	●	●	●	F2.2	●	●	●	F2.11	●	●	●	F4.1	●	×	●
F1.3	●	×	●	F2.3	●	×	●	F2.12	●	●	●	F4.2	●	○	×
F1.4	●	○	●	F2.4	●	×	●	F2.13	●	●	×	F4.3	●	○	○
F1.5	●	×	×	F2.5	●	×	●	F2.14	●	●	×	F4.4	●	○	○
F1.6	●	○	×	F2.6	●	○	●	F3.1	●	×	×	F4.5	●	●	●
F1.7	●	×	○	F2.7	●	○	●	F3.2	●	○	×	F4.6	●	×	×
F1.8	●	○	○	F2.8	●	●	●	F3.3	●	×	○				
F1.9	●	●	●	F2.9	●	●	●	F3.4	●	●	●				

●: same dependency, ○: different dependency, ×: the functionality is not included in the workflow

Table 5 Assessment of the suitability of the algorithms

Func	Algo	FEGS	TMS	CAS	FCS	Func	Algo	FEGS	TMS	CAS	FCS	Func	Algo	FEGS	TMS	CAS	FCS
F1.1	A1	●	○	×	○	F2.5	A18	●	×	×	×	F3.2	A31	●	●	×	×
F1.2	A2	●	●	●	×		A19	●	×	×	×		A32	●	●	●	×
	A3	●	●	○	○		A20	●	×	×	×		A33	●	×	×	×
F1.3	A4	●	●	×	×	F2.6	A21	●	×	×	×		A34	●	×	×	×
F1.4	A5	●	●	●	●		A22	●	×	×	×	F3.3	A35	●	●	×	●
F1.5	A6	●	○	×	×		A23	●	×	×	×		A36	●	○	×	○
F1.6	A4	●	●	●	×	F2.7	A24	●	●	×	×	F3.4	A37	●	●	●	●
F1.7	A7	●	●	×	●		A4	●	●	●	×		A38	●	●	●	●
F1.8	A8	●	●	●	●		A25	●	○	○	○	F3.5	A39	●	●	×	×
F1.9	A5	●	●	●	●	F2.8	A3	●	●	○	○	F4.1	A40	●	●	×	●
F2.1	A9	●	○	×	×		A5	●	●	●	●	F4.2	A41	●	○	×	×
	A10	●	○	×	×	F2.9	A9	●	○	×	×	F4.3	A42	●	○	○	○
	A11	●	×	×	×		A10	●	○	×	×	F4.4	A43	●	○	○	○
F2.2	A12	●	○	×	×	F2.10	A12	●	○	×	×	F4.5	A44	●	●	●	●
F2.3	A13	●	●	×	●	F2.11	A26	●	●	●	●	F4.6	A45	●	●	×	×
	A14	●	●	×	●	F2.12	A27	●	○	○	○						
F2.4	A15	●	×	×	×	F2.13	A28	●	○	○	×						
	A16	●	×	×	×	F2.14	A29	●	●	●	×						
	A17	●	×	×	×	F3.1	A30	●	●	×	×						

●: very suitable, ○: suitable after parametric modification, ×: not suitable sub-modules of the DCIP-MM, respectively. An overall view of the suitability of the algorithms to the target applications is shown in Table 5. It's worth mentioning that several algorithms were reused to realize different functionalities in the case of the fire evacuation guiding application, e.g. calculate the distance between two entities. Due to the diverse requirements of the target applications and the different attributes of the treated entities and situations, the qualitative evaluation had to consider three levels of fulfillment concerning the suitability of the algorithms, namely (i) very suitable, which meant that the implemented algorithms could be used directly in the target application, (ii) suitable with a parametric modification, which meant that the algorithms could be applied if some parameters were adjusted according to the requirements of the target applications (e.g. thresholds), and (iii) not suitable, which meant that the concerned algorithms were not suitable at all.

The result of the assessment of the suitability of the used data constructs is presented in Table 6. Two qualitative levels were considered in this assessment: (i) suitable, which meant that the data construct can sufficiently represented the needed content for information processing in the target application context, and (ii) not suitable, which meant that the data construct was either unneeded or insufficiently represented the required information. Incomplete data constructs did not fulfill the latter requirement.

3.4.3 Analysis of the empirical performance efficiency

Procedurally, this analysis involved (i) the specification of empirical performance indicators, (ii) the generation of the empirical performance profiles of the prototype in the I-CPS applications, and (iii) the exploration of the performance limitations in the case of the three I-CPS applications. In terms of the performance indicators, the work covered the efficiency analysis of the algorithms, as well as to the influencing factors on the actual performance. The results obtained from the performance tests in the fire evacuation guiding application were projected to the context of the three target application cases. Then, the computation time required by the algorithms in each of the target application cases were estimated and the practical characteristics of the target application cases, such as (i) the maximum possible number of entities handed at a time (the worst case) and (ii) the maximum sudden changes of entity numbers (the worst case). After this, the estimated computation time of algorithms was compared with the allowed computation time (*ACT*) in each of the application contexts. It was calculated by the following formula:

$$\mathrm{ACT} = \mathrm{ART} - (\mathrm{LTS} + \mathrm{LTM}) \tag{1}$$

where: *ART* is the allowed response time in a given application case, *LTS* and *LTM* are the practical latencies

Table 6 Assessment of the suitability of applied data constructs

Data	Used by algorithms	Applications			Data	Used by algorithms	Applications			Data	Used by algorithms	Applications		
		TMS	CAS	FCS			TMS	CAS	FCS			TMS	CAS	FCS
D1	A1	●	●	●	D18	A22	×	×	×	D35	A31	●	×	×
D2	A2, A4, A30, A31, A24, A35	●	●	×	D19	A23	×	×	×	D36	A34	●	×	×
D3	A2, A3	●	●	●	D20	A12	●	×	×	D37	A34	●	●	×
D4	A3	●	●	×	D21	A12, A24	●	×	×	D38	A34	×	×	×
D5	A4, A6	●	●	×	D22	A12	×	×	×	D39	A35	●	●	×
D6	A5	●	●	×	D23	A25, A43	●	●	●	D40	A36	●	×	●
D7	A5, A6, A8, A9, A10, A11, A24, A27, A32, A33, A34, A35, A36	●	●	×	D24	A4	●	●	●	D41	A39	●	×	●
D8	A7, A4	●	●	●	D25	A3	●	●	●	D42	A38	●	●	●
D9	A8	●	×	×	D26	A5	●	●	×	D43	A39	●	●	●
D10	A8	●	×	●	D27	A9, A10	●	●	×	D44	A42	●	●	●
D11	A12, A13, A14, A24	●	●	●	D28	A12	●	●	●	D45	A42, A43	●	●	×
D12	A27	●	●	●	D29	A26, A27, A30	●	●	●	D46	A44	●	●	●
D13	A15, A16, A17	●	●	●	D30	A37	●	●	●	D47	A44	●	●	●
D14	A18	●	×	×	D31	A28, A40	●	●	●	D48	A45	●	●	●
D15	A19	●	×	×	D32	A29, A41	●	×	×	D49	Output data	●	●	●
D16	A20	×	×	×	D33	A45, A34	●	●	●	D50	A41	●	●	×
D17	A21	×	×	×	D34	A33	●	●	●	D51	A42, A47	●	●	×

●: suitable, ×: not suitable

Table 7 Assessment of the efficiency of the algorithms in the target applications

Algo	TMS	CAS	FCS	Algo	TMS	CAS	FCS	Algo	TMS	CAS	FCS
A1	●	Δ	●	A16	Δ	Δ	Δ	A31	●	Δ	Δ
A2	×	●	Δ	A17	Δ	Δ	Δ	A32	●	●	Δ
A3	●	●	●	A18	Δ	Δ	Δ	A33	Δ	Δ	Δ
A4	×	●	Δ	A19	Δ	Δ	Δ	A34	Δ	Δ	Δ
A5	●	●	●	A20	Δ	Δ	Δ	A35	●	Δ	●
A6	●	Δ	Δ	A21	Δ	Δ	Δ	A36	●	Δ	●
A7	●	Δ	●	A22	Δ	Δ	Δ	A37	●	●	●
A8	●	●	●	A23	Δ	Δ	Δ	A38	●	●	●
A9	×	Δ	Δ	A24	×	Δ	Δ	A39	●	Δ	Δ
A10	●	Δ	Δ	A25	●	●	●	A40	●	Δ	●
A11	Δ	Δ	Δ	A26	●	●	●	A41	●	Δ	Δ
A12	●	Δ	Δ	A27	×	●	●	A42	×	●	●
A13	●	Δ	●	A28	●	●	Δ	A43	×	●	●
A14	●	Δ	●	A29	●	●	Δ	A44	●	●	●
A15	Δ	Δ	Δ	A30	●	Δ	Δ	A45	●	Δ	Δ

●: effective, ×: not effective, Δ: not needed

Table 8 Assessment of the effect of influencing factors on computation in the application cases

Influencing factors	Application cases			
	FEGS	TMS	CAS	FCS
Accuracy of positioning	High	Low	Low	High
Latency of information sensing	High	High	Low	High
Length of personalized messages	High	High	Low	High
Obedience of the informed user	Low	Low	High	High

of information sensing and messaging, respectively. If the estimated computation time of an algorithm was higher than the *ACT*, then the algorithm was regarded as 'not effective'. The results of the assessment are presented in Table 7.

Four factors influencing the practical performance of the system in the target application were considered. These included (i) the accuracy of positioning, (ii) the latency of information sensing, (iii) the length of personalized messages, and (iv) the obedience of the informed user. Because these factors strongly depend on the applied techniques for practical application cases, the assessments were made according to the characteristics shown in Table 1 and 2. The results of the investigation of the influential factors are summarized in Table 8.

3.4.4 Analysis of the theoretical performance sufficiency

While the specific performance indicators could be captured quantitatively, theoretical performance targets (TPTs) needed a qualitative evaluation due to their abstract and tentative nature. For the sake of comparability, we identified generic performance targets that were equally applicable to each target application. The TPTs chosen for the applicability analysis were:

- Adaptation need, which is judged by considering the total effort needed to adapt the elements of the modules (e.g. algorithms, data constructs) to the target application.
- Preparation effort, which is judged by considering the totaled preparation of a module (e.g. specification of thresholds in algorithms, and data constructs) when it is applied in the target application.
- Dependability, which is judged by considering the totaled dependency of the computational results generated by an adopted and prepared platform (all modules together) to the varying circumstances in the target application.

The research challenge was to demonstrate that in each of the three applications the TPTs could be achieved using the DCIP-MM or to show what sort of limitations has been experienced. The second part of the investigations focused on the sustainability of the overall performance of the CCMs under the influence of varying circumstances. The argumentation about the applicability of the DCIP-MM in the three application cases was based on the evidence obtained in the validation process with regards to quadrants one (relevance), two (suitability), and three (efficiency). The chosen TPTs were considered according to the following reasoning logic:

If either the testing of the functionality, or the testing of the overall workflow, or both, closes with a negative outcome, then the proposed mechanism cannot be applied in the target application cases, since the expectations for its theoretical structural relevance are not fulfilled.

If either the testing of the processing algorithms, or the testing of the processed data constructs, or both, concludes with a negative result, then the proposed mechanism cannot be applied in the target application cases, since its empirical structural suitability is not validated.

If either the specific performance indicators suggest poor efficiency or the factors influencing the computational efficiency have large effects on the outcomes (and make the computation unpredictable), or both concurrently appear, then the proposed mechanism cannot be applied in the target application cases.

If the theoretical structural relevance, the empirical structural suitability, and the empirical performance be all validated, then the sufficiency for an application depends on the achievement of the TPTs.

Based on the results discussed above in Sects. 4.1–4.3, items (a), (b), and (c) have been determined for the target application cases. Methodologically, three indicators have been specified for the assessment and assigned to the three qualitative TPTs, namely (i) theoretical relevance indicator (TRI), (ii) empirical suitability indicator (ESI), and (iii) empirical efficiency indicator (EEI). For each of these indicators, a percentage value was calculated, showing the proportion of applicable elements (e.g. function, algorithms, data constructs) of a mechanism or a module in a target application. For instance, if two algorithms and the related data constructs of a mechanism (or a module) were suitable for a target application, and if there were 10 algorithms in the mechanism in total, then we set the value of the ESI of the mechanism to 20% in the given target application. An overall applicability indicator (OAI) was defined, which could be evaluated based on the three individual indicators. The calculation was done by the following equation:

$$\mathrm{OAI} = \sqrt[3]{\mathrm{TRI} \times \mathrm{ESI} \times \mathrm{EEI}} \tag{2}$$

This equation establishes the geometric mean value of the three particular indicators capitalizing on the fact that they are interrelated. Therefore, the results of the percentile calculation concerning the applicability of the four modules (from M1 to M4) in the DCIP-MM are shown in Fig. 6.

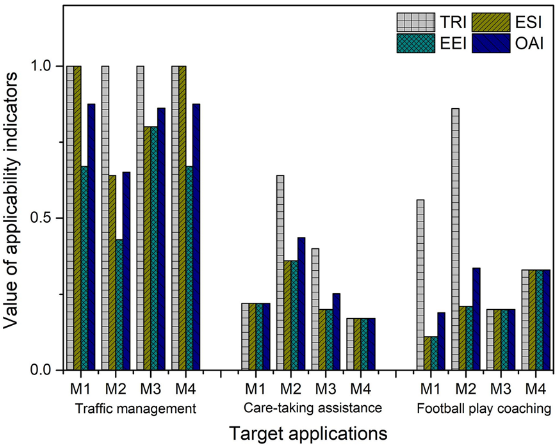


Fig. 6 Calculated indicators of the proposed mechanisms for the target applications

Including (i) high, (ii) medium, and (iii) low levels, three levels were introduced for the indicators of the TPTs, which were evaluated according to the principles shown in Table 9. The first TPT (adaptation need) was assessed based on the calculated OAI. The principle implies that if a low OAI is obtained, then a computational mechanism will need a large amount of adaptation work before it can be applied in the target application. The second TPT (preparation effort) was assessed only based on the calculated ESI. Our assumption is that the preparation effort for using a computational mechanism in a specific application depends on how much work is needed for changing the algorithms and data constructs. Based on the proposed principles, the results of assessment of the TPTs for the target application cases are shown in Table 10.

4 Discussion

In the tradition of scientific inquiry, internal validation and external validation goes hand-in-hand. A body of knowledge, which suffers from internal biases due to lack of procedural rigor, and/or the one, which is not appropriate for or has serious limitations with regards to its intended purpose, can be considered as scientific knowledge. Focusing on the usefulness of design methods, the original VSA targeted external validity, which manifests as the process of building confidence in the usefulness of a body of knowledge concerning a purpose. To this end, validation is driven by the question of whether the method provides design solutions 'correctly' (effectiveness), and whether it provides 'correct' design solutions (efficiency).

Table 9 Principles for assessing the TPTs

Assessed levels		Principles
Adaptation need		
	High	OAI ≤ 25%
	Medium	25% < OAI < 75%
	Low	OAI ≥ 75%
Preparation effort		
	High	ESI ≤ 25%
	Medium	25% < ESI < 75%
	Low	ESI ≥ 75%
Dependability		
	High	OSI ≥ 75%
	Medium	25% < OSI < 75%
	Low	OSI ≤ 75%

Table 10 Assessment results of the TPTs in the target application cases

Applications	TPTs	M1	M2	M3	M4	Overall
Traffic management	Adaptation need	Low	Medium	Low	Low	Low
	Preparation effort	Low	Medium	Low	Low	Low
	Dependability	–	–	–	–	Medium
Care-taking assistance	Adaptation need	High	Medium	Medium	High	Medium
	Preparation effort	High	Medium	High	High	High
	Dependability	–	–	–	–	Low
Football-play coaching	Adaptation need	High	Medium	High	Medium	Medium
	Preparation effort	High	High	High	Medium	High
	Dependability	–	–	–	–	High

In this context, correct design solutions provide acceptable operational performance (e.g. are designed and realized with less cost and/or in less time). This needs specific measures in the process of theory generation and design methodology development, as well as rigorous testing when the result of the process is available. In addition, the VSA also recognized that ‘formal, rigorous and quantitative’ validation cannot be applied without problems in certain areas of engineering research, which rely more on subjective statements than on physical experimentation and mathematical modeling.

The above assumptions have been reused in the A-VSA, which incorporated the criteria for validation of the applicability of constructive computational mechanisms (CCMs). As indicated by the reasoning model, applicability validation of CCMs relies both on subjective statements and on mathematical modeling. For this reason, the A-VSA combines both qualitative and quantitative perspectives, as well as theoretical and empirical viewpoints. Using the A-VSA, the applicability of CCMs can be assessed in terms of four aspects (i) theoretical structural validity, (ii) empirical structural validity, (iii) empirical performance validity, and (iv) theoretical performance validity. The measure of theoretical structural appropriateness is relevance. It is evaluated by considering whether (i) the offered functionality is relevant for the target application, and (ii) the overall computational workflow is relevant for the target application. The empirical structural appropriateness is expressed as suitability. It is evaluated by considering if (i) the processing algorithms are suitable for the target application, and (ii) the processed data constructs are suitable for the target application. The empirical performance is measured in terms of efficiency, which expresses if (i) the specific performance indicators confirm efficiency, and (ii) the effects of the factors influencing the processing are controllable. Finally, the theoretical performance is expressed in terms of the overall sufficiency, which is characterized by (i) the extent of achieving the overall performance targets, and (ii) the level of sustainability of the overall performance in varying circumstances.

In the conducted case study, three completely different application cases were used as the basis of the applicability evaluation of the already developed CCM, namely, the DCIP-MM. Both the theoretical validity and the empirical structural validity were evaluated in the assessment procedure. The qualitative assessment concerned the measures of relevance and suitability to confirm the internal appropriateness of the DCIP-MM in the target applications. On the other hand, quantitative evaluations were applied to quantify the performance validity based on various specified measures. It’s worth mentioning that the concerned measures were dedicated to the DCIP-MM. For instance, we selected the computation time as a measure of empirical performance, since it is one of the most crucial aspects of dynamic context information processing. Nevertheless, many other quantitative measures can be applied when using the A-VSA to other CCMs, such as throughput, time, and space complexity. Concerning quadrants 1–4, several indicators (i.e. TRI, ESI, EEI, and OAI) were used to show the percentage of the elements of a CCM applicable to target applications. These indicators jointly contribute to a qualitative evaluation of the theoretical performance targets. Concerning the calculation process, we only considered the proportion of available elements in each module (number of reusable elements over the total number of elements) to indicate the applicability. However, the calculations can be further refined depending on the adaptation effort concerning the algorithms and/or functionalities in the target applications. The research confirmed that qualitative applicability evaluation could be transferred into a quantitative assessment when concrete requirements were specified for the four quadrants of A-VSA. Accordingly, a decision can be made concerning how much a mechanism can match the applicability expectations.

Though applicability testing focused on the external validity (that is, the appropriateness of applying a mechanism), the evaluation process had to confirm the internal validity of a CCM (checking the internal appropriateness in target application) first. It was observed in the evaluation process that there was a tight coupling among the four concerned aspects of evaluation, which were associated with certain dependencies. If the computational mechanism fails in any one of the concerned aspects, then the theoretical performance targets (sufficiency) cannot be fulfilled and there is no way to conclude about the applicability of the DCIP-MM positively in the considered application. If the relevance and suitability are positively evaluated, then the empirical performance plays a decisive role in the judgment. In those cases, where the relevance and suitability indicators present a partial compliance only, and the efficiency designate a partial completion, the functional structure, the procedural flow, the algorithms and data structures, and the computational measures of the DCIP-MM should be adapted to those required by a specific application.

The obtained results cast light on two facts. On the one hand, the theory and approach proposed in the paper are efficient and appropriate for the applicability validation of CCMs. On the other hand, the significant differences in terms of the case characteristics of the chosen target applications define a reasonably broad spectrum of applications. Having a narrower or broader range of the case characteristics, many other applications with rather different purposes have the potential to be involved.

5 Conclusions

The objectives of this paper were to clarify the concepts and criteria for applicability validation of constructive computational mechanisms (CCMs) and to propose an approach that can support this. Towards this end, evidence about measures and/or indicators of applicability has been obtained with regards to (i) functional/procedural relevance, (ii) suitability of algorithms and data constructs, (iii) efficiency of specific performances, and (iv) sufficiency of overall performance. An adapted validation square approach (A-VSA) has been developed for assessing the external appropriateness and usefulness of CCMs from both theoretical and empirical points of view. As evidenced by the completed applicability-testing cross-case study, the developed A-VSA is an efficient tool for assessing the external validity of CCMs, and the proposed measures and indicators are adequate and expressive enough. The results of applicability testing can guide the adaptation work of a CCM when applying it to an out-of-domain application.

Based on the conducted research, we could conclude that applicability validation of CCMs required both subjective statements and mathematical modeling. This is in concert with the tradition of validation in engineering research, that is anchored in logical induction and/or deduction, whereas most existing validation approaches for CCMs only address physical experimentations and mathematical modeling. The advantage of the A-VSA is that it offers a way to transfer qualitative applicability evaluation into quantitative applicability assessment, which enables the use of both subjective statements and mathematical modeling in the course of applicability testing. In addition, the research outcome indicates that the external validation of a method depends on the results of its internal validation. The external validity of CCMs cannot be approved before the internal validity is confirmed. Furthermore, the A-VSA establishes a tight coupling among the enablers embraced by a CCM and the aspects of theoretical and empirical validation. Hence, the proposed A-VSA is not only an efficient tool for defining the range and/or the extent of applicability of various CCMs, but it can also be used to explore potential applications, and even to create new applications for existing CCMs.

Acknowledgements The work is supported by National Natural Science Foundation of China (52105323), Natural Science Foundation of Jiangsu Province (BK20210232), and Fundamental Research Funds for the Central Universities (2242021R10022).

Author contributions Conceptualization: Yongzhe Li and Imre Horváth; Methodology: Imre Horváth and Zoltán Rusák; Formal analysis and investigation: Yongzhe Li and Imre Horváth; Writing—original draft preparation: Yongzhe Li; Writing—review and editing: Imre Horváth and Zoltán Rusák; Funding acquisition: Yongzhe Li; Supervision: Imre Horváth.

Funding The work is supported by National Natural Science Foundation of China (52105323), Natural Science Foundation of Jiangsu Province (BK20210232), and Fundamental Research Funds for the Central Universities (2242021R10022).

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Ahmad W, Qamar U, Hassan S (2015) Analyzing different validation and verification techniques for safety critical software systems. Proceedings of the 6th IEEE international conference on software engineering and service science, Beijing, China, pp 367–370. https://doi.org/10.1109/ICSESS.2015.7339076

Barambones J, Imbert R, Moral C (2021) Applicability of multi-agent systems and constrained reasoning for sensor-based distributed scenarios: A systematic mapping study on dynamic DCOPs. Sensors (basel, Switzerland) 21(11):3807. https://doi.org/10.3390/s21113807

Barlas Y, Carpenter S (1990) Philosophical roots of model validation: two paradigms. Syst Dyn Rev 6(2):148–166. https://doi.org/10.1002/sdr.4260060203

Brazdil P, Gama J, Henery B (1994) Characterizing the applicability of classification algorithms using meta-level learning. In: Bergadano F., De Raedt L. (eds) Machine Learning: ECML-94. Lecture

Notes in Computer Science 784. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-57868-4_52

Brings J, Bohn P, Bandyszak T, Föcker F, Daun M (2016) Model-based prototype development to support early validation of cyber-physical system specifications. Second international workshop on requirements engineering for self-adaptive and cyber-physical systems, Gothenburg, Sweden, pp 1–15

Christophe F, Bernard A, Coatanéa E (2010) RFBS: a model for knowledge representation of conceptual design. CIRP Ann Manuf Technol 59(1):155–158. https://doi.org/10.1016/j.cirp.2010.03.105

Dauby JP, Dagli CH (2011) The canonical decomposition fuzzy comparative methodology for assessing architectures. IEEE Syst J 5(2):244–255. https://doi.org/10.1109/JSYST.2011.2125250

Debbabi M, Hassaïne F, Jarraya Y, Soeanu A, Alawneh L (2010) Verification and validation in systems engineering - Assessing UML/SysML design models. Springer, Verlag Berlin Heidelberg

Donald JG (1995) Disciplinary differences in knowledge validation. New Dir Teach Learn 64:7–17

Du Bois E, Horváth I (2013) Operationalization of the quadrant-based validation in case of a designerly software development methodology. Proceedings of the 19th international conference on engineering design, pp 365–376

Eze T, Anthony RJ, Walshaw C, Soper A (2011) The challenge of validation for autonomic and self-managing systems. Proceedings of the seventh international conference on autonomic and autonomous systems, pp 128–133.

Feth P, Bauer T, Kuhn T (2015) Virtual validation of cyber physical systems. In: Proceedings of the software-engineering and management, Bonn, Germany, pp 201–206.

Frey DD, Dym CL (2006) Validation of design methods: lessons from medicine. Res Eng Des 17(1):45–57. https://doi.org/10.1007/s00163-006-0016-4

Fu K, Murphy J, Yang M, Otto K, Jensen D, Wood K (2015) Design-by-analogy: experimental evaluation of a functional analogy search methodology for concept generation improvement. Res Eng Design 26:77–95. https://doi.org/10.1007/s00163-014-0186-4

Gonzalez AJ, Barr V (2000) Validation and verification of intelligent systems-What are they and how are they different? J Exp Theor Artif in 12(4):407–420. https://doi.org/10.1080/095281300454793

Guarro S, Ozguner U, Aldemir T, Knudson M, Kurt A, Yau M, Hejase M, Kwon S (2016) Formal validation and verification framework for model-based and adaptive control systems. In: Rayadurgam S, Tkachuk O (eds) NASA formal methods. Springer, Cham, pp 227–233. https://doi.org/10.1007/978-3-319-40648-0_18

Horváth I (2020) Sympérasmology: a proposal for the theory of synthetic system knowledge. Designs 4(47):1–24. https://doi.org/10.3390/designs4040047

Horváth I, Li Y, Rusák Z, van der Vegte WF, Zhang G (2017) Dynamic computation of time-varying spatial contexts. J Comput Inf Sci Eng 17(1):011007. https://doi.org/10.1115/1.4034034

Kroll E, Weisbrod G (2020) Testing and evaluating the applicability and effectiveness of the new idea-configuration-evaluation (ICE) method of conceptual design. Res Eng Des 31:103–122. https://doi.org/10.1007/s00163-019-00324-6

Li Y, Horváth I, Rusák Z (2019) Personalized messaging based on dynamic context assessment: Application in an informing cyber-physical system. J Integr Des Process Sci 23(1):103–123. https://doi.org/10.3233/JID190004

Li Y (2019) Utilizing dynamic context semantics in smart behavior of informing cyber-physical systems. Dissertation, Delft University of Technology

Mejía-Gutiérrez R, Carvajal-Arango R (2017) Design verification through virtual prototyping techniques based on systems engineering. Res Eng Des 28(4):477–494. https://doi.org/10.1007/s00163-016-0247-y

Olewnik AT, Lewis K (2005) On validating engineering design decision support tools. Concurrent Eng 13(2):111–122. https://doi.org/10.1177/1063293X05053796

Pape L, Giammarco K, Colombi J, Dagli C, Kilicay-Ergin N, Rebovich G (2013) A fuzzy evaluation method for system of systems meta-architectures. Procedia Comput Sci 16:245–254. https://doi.org/10.1016/j.procs.2013.01.026

Pardo SA (2016) Validation and verification. In: Empirical modeling and data analysis for engineers and applied scientists. Springer, Cham, pp 197–201. https://doi.org/10.1007/978-3-319-32768-6_13

Parkes DC (2008) Computational Mechanism Design. Lecture notes of tutorials presented at the 10th Conference on Theoretical Aspects of Rationality and Knowledge TARK-05, Singapore.

Pedersen K, Emblemsvåg J, Bailey R, Allen JK, Mistree F (2000) Validating design methods & research: The validation square. Proceedings of the ASME design engineering technical conferences, DETC2000/DTM-14579, Baltimore, Maryland, pp 1–13.

Reich Y (2017) What is a reference? Res Eng Des 28(4):411–419. https://doi.org/10.1007/s00163-017-0270-7

Schlesinger S (1979) Terminology for model credibility. Simul 32(3):103–104. https://doi.org/10.1177/003754977903200304

Seepersad CC, Pedersen K, Emblemsvåg J, Bailey RR, Allen JK, Mistree F (2006) The validation square: how does one validate design methods?" Decision-Based Design: Making Effective Decisions in Product and Systems Design, New York, ASME Press, Chapter 25, pp. 305-326.

Smolka J, Bíró O, Nowak AJ (2009) Numerical simulation and experimental validation of coupled flow, heat transfer and electromagnetic problems in electrical transformers. Arch Comput Methods Eng 16:319–355. https://doi.org/10.1007/s11831-009-9032-7

Teegavarapu S (2009) Foundations of design method development. Dissertation, Clemson University

Zhao M, Lu H, Yang S, Guo F (2020) The experience-memory Q-learning algorithm for robot path planning in unknown environment. IEEE Access 8:47824–47844. https://doi.org/10.1109/ACCESS.2020.2978077

Zheng C, Hehenberger P, Duigou JL, Bricogne M, Eynard B (2016) Multidisciplinary design methodology for mechatronic systems based on interface model. Res Eng Des 28(3):333–356. https://doi.org/10.1007/s00163-016-0243-2