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
Accepted author manuscript

Published in
Research in Engineering Design

(Manuscript for Research in Engineering Design)

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

Yongzhe Li [a,b], Imre Horváth[b], Zoltán Rusák[b]

a. School of Mechanical Engineering, Southeast University, Nanjing, P. R. China

b. Faculty of Industrial Design Engineering, Delft University of Technology, Delft, the Netherlands

Corresponding author: Yongzhe Li

E-mail address: yongzheli@seu.edu.cn

Declarations

Funding:

The work is supported by "the Fundamental Research Funds for the Central Universities".

Conflicts of interest/Competing interests:

The authors declare no conflict of interest.

Availability of data and material:

Not applicable

Code availability:

Not applicable

Authors' 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

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

Yongzhe Li [a,b], Imre Horváth[b], Zoltán Rusák[b]

a. School of Mechanical Engineering, Southeast University, Nanjing, People's Republic of China

b. Faculty of Industrial Design Engineering, Delft University of Technology, Delft, the Netherlands

Abstract: Applicability testing of constructive computational mechanisms (CCMs) is a new challenge for both academia and industry, for the reason that many CCMs are developed for a given application. A typical example is an automated car parking CCM. In this paper, we propose the adapted validation square approach (A-VSA) for a systematic evaluation of the applicability of a CCM to cases that were not considered during its design time. The A-VSA makes it possible to evaluate its appropriateness from (a) theoretical structural, (b) empirical structural, (c) theoretical performance, and (d) empirical performance dimensions. Altogether eight aspects were considered in the evaluation process, for instance, relevance of functionalities and suitability of data constructs. As test case for the validation of the A-VSA approach, a dynamic context information processing multi-mechanism (DCIP-MM) was considered. This computational mechanism was developed for a conceptualized indoor fire evacuation guiding system. The paper explains the applicability evaluation strategy of the A-VSA, and operationalizes it in the context of application of the DCIP-MM in three completely different application cases, which represent the presumed boundaries of applicability of this complex CCM. It has been concluded that the A-VSA is an efficient methodology for applicability validation of CCMs. The advantage of the A-VSA is that the indicators can be replaced by concrete requirements and thereby the qualitative applicability evaluation can be transferred into a quantitative applicability assessment.

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

1. Introduction

1.1. Introducing the addressed research issue

Constructive computational mechanisms (CCMs) are purposefully designed and implemented algorithmic structures to support various knowledge-intensive activities (Sipser 2006). As main forms of knowledge inferring/reasoning, both non-ampliative and ampliative mechanisms have been proposed over the last decades. Application-specific reasoning mechanisms, which attempt achieving a balance between application neutrality/specificity and performance dependability/efficiency, have also been considered (Horváth 2020). Currently, there is a paradoxical situation concerning their development. The input information needed for development is typically obtained by considering only one or a limited number of 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). 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 the academia and the industry.

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’ by 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 are necessary and useful. In the tradition of scientific inquiry, validation is a multi-faceted activity focusing on confirmation of knowledge. As discussed by Barlas and Carpenter (1990), while verification refers to internal consistency, validation refers to 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 fulfilment 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 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 on 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, development of applicability validation methodologies typically considers a limited number of application cases only, whereas the developed methodology is supposed to cover the broadest possible range of application cases. This paper intended to resolve this paradoxical situation, starting out from a theoretically already well established and a methodologically reasonably transparent validation approach. The validation square approach, originally considered for validation of design methods is a promising starting point towards an approach dedicated to applicability testing (Pedersen et al. 2000). Therefore, this approach 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).

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 space limitation, only the contributions to the subject area of validation of methods, methodologies and tools were summarized, where constructive computational mechanisms 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 (2019) 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 four types of prototypes that are used in design validation during product development, namely (i) abstract prototypes, (ii) virtual prototypes, (iii) functional prototypes, and (iv) physical prototype.

As a kind of articulated method that addresses on particular computational purposes, constructive computational mechanisms require systematic approaches to approve its validity. Eze et al. (2011) discussed various challenges associated with validation of self-managing and autonomic computational mechanisms, and investigated the relations of validation approaches (such as system unit testing, real-world system testing, pervasive supervision, system model checking, system self-testing) and execution modes (generic, design-time, run-time, integrated, autonomous). Ahmad et al. (2015) analyzed validation techniques for safety-critical software such as (i) functional failure analysis, (ii) hazard and operability studies, (iii) failure modes and effects analysis), and (iv) fault trees analysis, and compared them 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 based on model-based prototype development.

Gonzalez and Barr (2000) exposed the differences to be taken into account 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 validation of design decision methodologies. According to them, the complexity of prescriptive models makes their validation a difficult task.

Complex systems usually comprise of a large number of interacting modules, which requires 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/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 for the purpose of verification and validation of mechatronic systems. Dauby and Dagli (2011) proposed a methodology for the assessment of 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 approach should be logically rigorous, internally consistent, and mathematically correct. However, the literature offers only limited insights with regards to the validation of constructive computational methods. This is because the validation is largely influenced by the purpose of their applications, which addresses on internal validity. Accordingly, external appropriateness should also be focused on. The typical aspects of validation in the concerned external appropriateness should be checked, for instance, in terms of usability, applicability, performance and overall utility.

1.4. The focus and contents of this paper

This paper proposes an approach for applicability validation of constructive computational methodologies, and uses the approach on a computational mechanism as a case study. The next section discusses the essence and features of both the original VSA and the adapted validation square approach (A-VSA). Section 3 presents a dynamic context information processing multi-mechanism (DCIP-MM) and potential applications that the mechanism is targeting. Section 4 focuses on the details about using the A-VSA to evaluate the applicability of the DCIP-MM. Conclusions and future research are given in Section 5.

2. Towards operationalization of the adapted validation square approach

2.1. The validation square approach and its features

The validation square approach (VSA), initially proposed by Pedersen et al. (2000) is a very powerful approach for external validation of design methods. The reasoning model of the VSA is shown in Figure 1. 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:

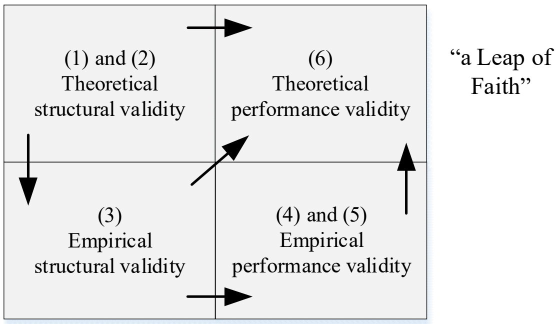

Fig. 1 The reasoning model of the validation square approach (VSA)

(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;

(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 finally

(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 requires a 'leap of faith', which can be created in steps from (i) to (iii). They make it possible to argue about domain-specific performance validity, if the method is useful in a specific sense, and about domain-independent performance validity, if the method is found useful in a more general sense.

It has been mentioned as a limitation of the VSA that it does not offer specific methods/tools of validation (Teegavarapu 2019). Furthermore, 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. Tailoring the VSA to applicability testing

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 with respect to an application purpose. As a criterion of applicability, appropriateness indicates if the concerned CCM is able to produce utility, and usefulness indicates that it can enable the embedding system work correctly. The latter criterion was captured in VSA as 'effectiveness', and the former criterion as 'efficiency'. Effectiveness was supposed to be assessed by a qualitative evaluation, while efficiency by quantitative assessment. 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. Figure 2 shows the fundamental questions that expressed the considered technical measures of relevance, suitability, efficiency and sufficiency.

In the process of evaluation, the A-VSA considers: (i) the relevance of the structural constituents of the DCIP-MM, (ii) the sufficiency of the theoretical performance targets and criteria, (iii) the suitability of the empirical procedural/information flow and data constructs, and (iv) the efficiency of the empirical performance enablers as indicators and measures of applicability. If the match of the theoretical structure of the constituents of the DCIP-MM cannot be shown, then it cannot be applied to the particular case. However, the theoretical structural validity is only a necessary condition, but not a sufficient condition of applicability. The sufficiency criterion of structural appropriateness can be met if the algorithms and the logical and functional dependences between them, as well as the data, information and knowledge constructs/structures processed by the algorithms are appropriate for the particular application. Likewise, the efficiency of the empirical performance is only a necessary condition for usefulness.

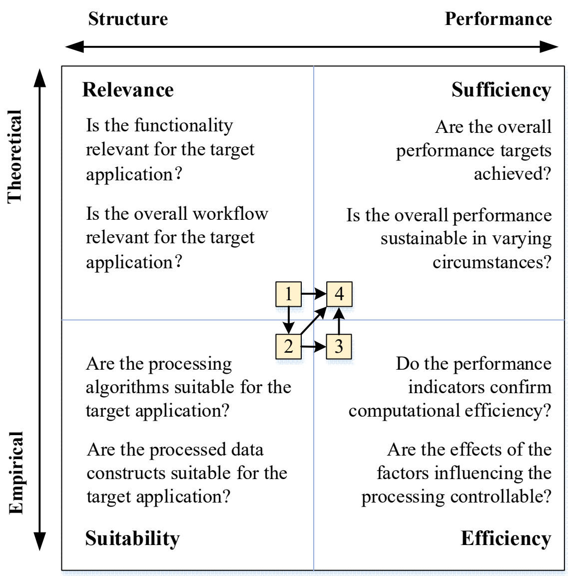

Fig. 2 The strategy of applicability validation using the adapted validation square approach (A-VSA)

The sufficient condition is that all theoretical performance target should be achieved. The 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 applicability characterization in real-life application cases. On the other hand, they are also supposed to hint at necessary or possible adaptations towards a high-level appropriateness and usefulness. The indicators can be chosen so as 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.

Evaluation of the empirical performance needs a strategy to identify and rank the critical factors and procedures based on their influences and impacts. Empirical computational performance of the mechanisms and algorithms, and the computability, reliability and completeness of the data, information and knowledge structures/constructs can be expressed by measures such as (i) amount of input data, (ii) time of data preparation, (iii) simplicity of models, (iv) input efforts, (v) computation time, (vi) amount of output data, (vii) sensitivity of the outcome to errors, etc. can be considered as concrete quantitative indicators of efficiency. In addition, various non-numerical indicators can also be considered such as (i) ease of use, (ii) human cognitive effort, (iii) trust building by individuals, (iv) chance of making errors, (v) transparence of inputs/outputs, (vi) awareness of the work of computational mechanisms/algorithms, etc. These typically need qualitative description or characterization. Performance deficiencies can be caused by factors such as (i) availability of cloud resources, (ii) malfunctioning of algorithms, (iii) logical fallacies in reasoning, (iv) temporal incompleteness or incorrectness of data, (v) sensitivity of data/computation to deviations, etc. The theoretical performance targets can be such as (i) cost reduction, (ii) sparing time, (iii) reducing efforts, (iv) eliminating errors, (v) resolving bottlenecks, (vi) complementing knowledge, (vii) facilitating collaborations, (viii) achieving innovation, and (ix) improving quality. This kind of targets can be used in application cases other than the applications, which provided input requirements and information for the conceptualization and development of a particular mechanism. This enhances the external validation flavor of the A-VSA.

3. A case study: Validation of the A-VSA approach

3.1 Introducing a developed CCM

To support dynamic context information processing in informing cyber-physical systems (I-CPSs), a dynamic context information processing multi-mechanism (DCIP-MM) has been developed, which support various knowledge-intensive decision making and action planning activities. The DCIP-MM includes four modules/sub-mechanisms, which are dedicated to (i) dynamic context information representation, (ii) semantic context knowledge information inferring, (iii) context-based real-life action plan deriving, and (iv) context-sensitive message construction and delivery. The overall functionalities of the mechanisms included in the DCIP-MM are shown in Figure 3.

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. Each of the interoperating mechanisms comprises several functionalities and algorithms. It was also a design principle that the DCIP-MM as a software platform can be integrated and can interoperate with other software constituents of I-CPSs. Further information about the theoretical fundamentals, conceptualization and implementation of the DCIP-MM can be found in (Li 2019).

The conceptualization of the DCIP-MM considered specific requirements for a fire evacuation

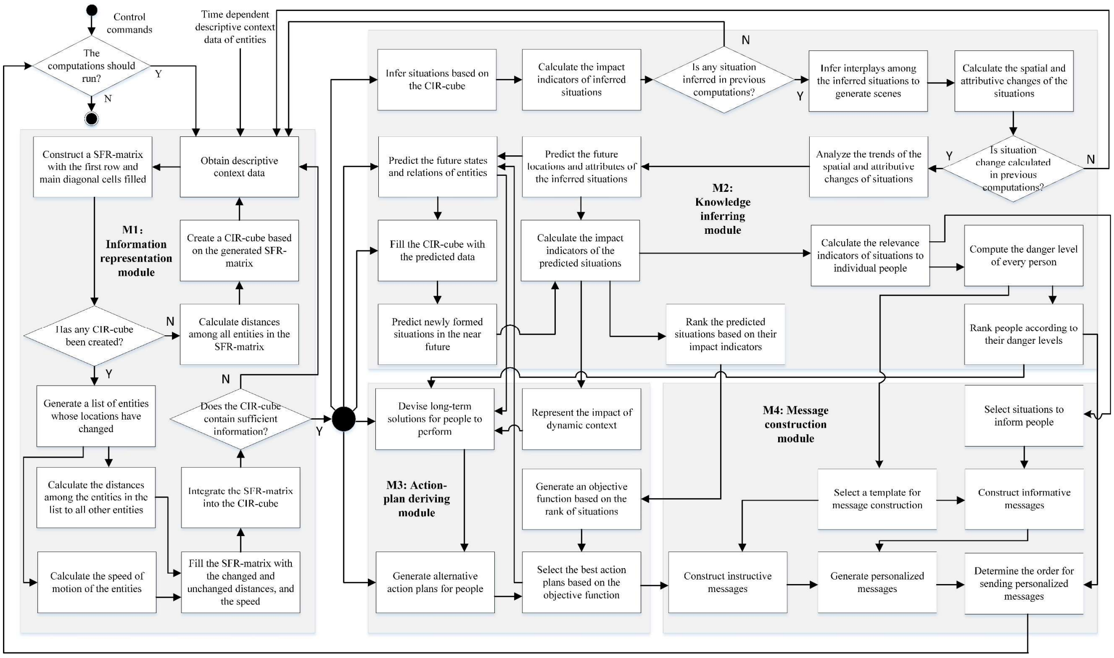

Fig. 3 High-level overview of the specific functionalities included in a DCIP-MM

guiding application. 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 methodology.

3.2. The reference application case

The proposed DCIP-MM has been developed with an indoor fire evacuation guiding (FEG) application in mind. It was used both as a computational scenario and as a source of application data/knowledge. A symbolic image of the FEG is shown in Figure 4. The FEG system aims at providing personalized escape strategy for people in different situations. The main research 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

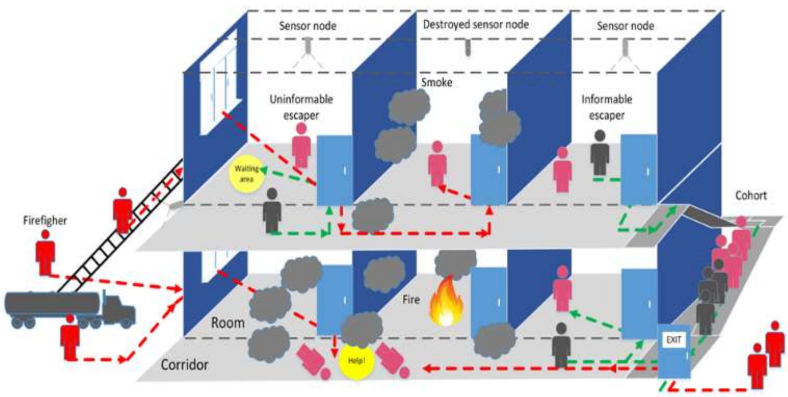

Fig. 4 The reference application case: indoor fire evacuation guiding (FEG) application

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 in order 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 is able to provide both descriptive and predictive knowledge about emergency situations as well as about the implications of the interplaying situations on the entities in quasi-real time. Nevertheless, the DCIP-MM have been designed with the objective of reusability in comparable target applications. This raises the need for investigation and validation of its applicability in the considerable other application cases.

3.3. The chosen target application cases

Called target applications, three possible but not trivial I-CPS applications have been considered. Illustrative images of these applications are shown in sub-figures 5.a – 5.c. Each of them needs dynamic

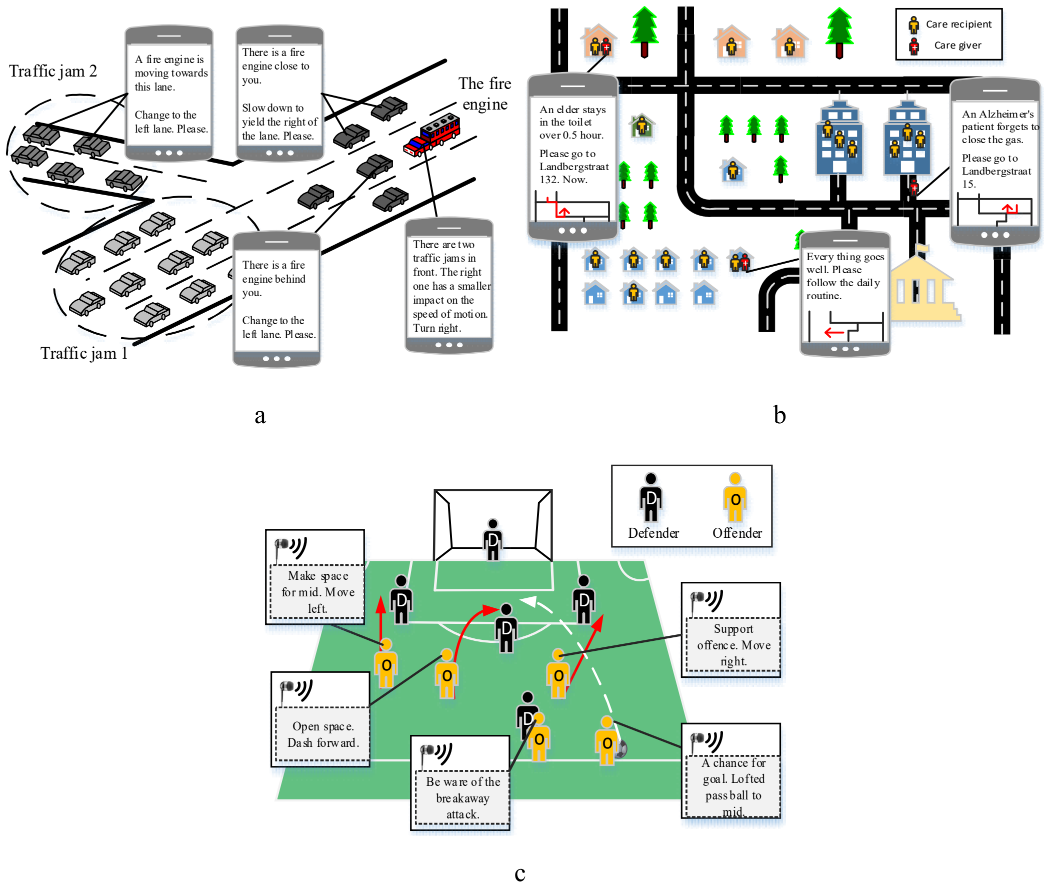

Fig. 5 The target application cases of the DCIP-MM: (a) traffic management system (TMS), (b) care taking assistance system (CAS), and (c) football-play coaching system (FCS)

context information processing, situated reasoning, and personalized message communication, which lend themselves to novel applications of a smart I-CPS.

Application 1: The traffic management system (TMS) is deemed to be necessary and useful due to the increase of the number of vehicles on city roads. As shown in Figure 5.a, a specific scenario considered for traffic management is given as follows. A fire engine is moving on road for a fire distinguishing mission. There are two traffic jams accidentally formed on the high-way, which may influence the movement of the fire engine. In this case, the objective of the TMS is to determine the optimal routes and actions for the related drivers to enable the fire engine arrives 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. The dynamically changing locations, attributes and relations together form different situations may also influence the decision-making process of the TMS.

To manage the different 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 concerning (a) the essence of situations, (b) the implications of the situations on the concerned entities, and (c) the relevance of the situations to these 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.

Application 2: A home caretaking assistant system (CAS) provides caretaker services for care recipients (e.g. elderlies, patients, handicapped, etc.) in a home environment. A specific scenario for the application is shown in Figure 5.b. Three care givers provide daily care for a group of people. As their routine tasks, each of the care givers 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. They may need to look after a patient who is not in the daily schedule or to change the priority concerning the care recipients. Furthermore, if more than one dangerous situation happens at the same time, then they have to make decisions about a plan of actions. To support the care givers' work, the CAS should be involved in monitoring, manage the schedule, reason with the event and the servicing capacity of the care givers, and provide action plans and information in a real-time manner.

To achieve this, the dynamic context information of caregivers and care recipients should be processed. The dangerous situations in which the care recipients may be involved by can be inferred based on various pieces of information aggregated in their home. The challenges are (i) to prioritize the care recipients based on the inferred situations and the distances to the caregivers, and (ii) to manage the spatiotemporal context of the care givers (since every action takes time to get completed). Therefore, the CAS is supposed (i) to aggregate information about the personal context of care recipients (e.g. activities) as well as about the care giver (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 the care of care recipients.

Application 3: The third case is a real-time football-play coaching system (FCS). Obviously, training is indispensable to enhance the performance quality and to increase the efficiency of playing football. The system is supposed to real-time contact with players of a team simultaneously to improve group performance. Figure 5.c shows a hypothetical scenario. The offenders shown are given with 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) during the training sessions. 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 context. The FCS should (i) develop strategies for completing the offense according to the dynamic context changes, (ii) generate action instructions based on the spatial position, attributes and role of the players, and (iii) instruct the players to replay the actions in the given situation.

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. The dynamic context of the individual players is decisive in terms of: (i) to which location to move on the playground, (ii) what speed to take to reach the target location, and (iii) what actions to take after receiving the ball. To advice players about the 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).

3.4. Comparison of the concerned target applications

The above description of the three target I-CPS application cases casts light on two facts. One the one hand, there are significant differences in terms of the case characteristics of the chosen target applications, which in turn define a reasonably broad spectrum of applications. Having a narrower or broader range of the case characteristics, many other applications with rather different purposes can be involved. On the other hand, enabling technologies applied by the chosen target applications may involve different constraints on computation, e.g. communication and computation hardware. The realization of expected performance should depend on the empirical influencing factors. Therefore, the characteristics and influencing factors of the chosen target applications were focused, in order to create a factual basis for the applicability analysis and validation.

We selected three major characteristics for the purpose of comparative applicability analysis, 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 computations. The specific

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	Moderate (40-300)	Large (100-1000)	Moderate (20-100)	Small (23)
Required response time	Moderate (< 5s)	Short (< 1s)	Long (< 10s)	Short (< 0.5s)
The sudden change of entity number	Moderate (0-10)	Large (10-200)	Small (0)	Small (0)

Table 2: A comparison of the technical factors that influence the empirical performance of the target I-CPS applications

Influencing factors	Application cases			
	FEGS	TMS	CAS	FCS
Used sensing technology	UWB	GPS	GPS + Indoor camera	Camera
Typical positioning accuracy	Moderate (< 0.5 m)	Low (< 4.9 m)	Low (< 4.9 m)	High (< 0.1 m)
Latency of information sensing	Moderate (< 0.25 s)	Long (< 0.5 s)	Long (< 0.5 s)	Short (< 0.1 s)
Power supply of terminals	Moderate (hand-held)	High (on-board)	Moderate(hand-held)	Low (in-ear)
Technology for message sending	WLAN/WiFi (10Mbps)	4G Mobile web (100 Mbps)	4G Mobile web (100 Mbps)	Bluetooth 5.0 (2Mbps)
Latency of message receiving	Moderate (5-20 ms)	Long (60-100 ms)	Long (60-100 ms)	Short (<3 ms)
Length of personalized messages	Moderate (10-30 words)	Moderate (10-30 words)	Large (50-80 words)	Short (0-10 words)
Latency of message reading	Long (0-∞ s)	Short (0 s)	Long (0-∞ s)	Short (0 s)

quantitative values associated with these characteristics are indicated in Table 1. Although there might be some exception situations, we assumed that these values could properly fit most cases in the application contexts. As shown in Table 2, eight factors influencing information sensing and messaging were considered to represent the technical constraints in the chosen application cases. The technological and operational comparison shows that the empirical characteristics and the concerned aspects of influencing factors are rather varied in the selected target application cases. Nevertheless, each of these applications can be considered as a representative of a family of slightly different application cases sharing similar characteristics and situations.

4. Validation of the applicability of the DCIP-MM to the target applications

4.1. Analysis of the relevance of the theoretical structure and the workflow of the DCIP-MM

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 mechanisms and to the relevance of the processing workflow of the DCIP-MM to the traffic management, care-taking assistance and football-play coaching 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 means that the functionality is fully required by the target application, (ii) partially relevant, which means that the functionality can partially fulfil the requirement of the target application, and (iii) not relevant, which means that the functionality is not required by the target application at all.

To evaluate the relevance of workflow, functionality dependency was used to characterize the relationship of functionalities in a workflow. For a given functionality that is needed by an application case, its previously performed functionalities define the dependency. Accordingly, a comparison of the dependency of the functionalities for the concerned applications is shown Table 4. The assessment considered three values: (i) same dependency, which means that the functionalities performed before a concerned one in the target application is the same as those specified in the DCIP-MM, (ii) different dependency, and (iii) the functionality is not included in the workflow. Actually, the dependency of the functionalities for traffic management is the same as in the case of the fire evacuation guiding. It means that the workflow can directly be reused. For the CAS and the FCS, some functionalities specified for the fire evacuation are not needed, which negatively influence the reusability of the workflow. For details about the assessment process, please refer to (Li 2019).

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

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

●: Very relevant, ○: Partially relevant, ×: Not relevant

4.2. Analysis of the suitability of the empirical structure of the DCIP-MM

Every functionality should be realized through dedicated algorithms and appropriate 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 the traffic management, care-taking assistance and football-play coaching applications. There are 8, 21, 10 and 6 algorithms in the sub-modules of the DCIP-MM, respectively. An overall view of the suitability of the algorithms to the targets applications is shown in Table 5. It's worth mentioning that several algorithms were reused to realize different functionalities in the case of 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 take three levels of fulfilment into consideration with regard to the suitability of the algorithms, namely (i) very suitable, which means that the implemented algorithms can directly be used in the target application, (ii) suitable with a parametric modification, which means that the algorithms can be applied if some parameters are adjusted

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

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

according to the requirements of the target applications (e.g. thresholds), and (iii) not suitable, which means that the concerned algorithms are 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 means that the data construct is able to sufficiently represent the needed content for information processing in the target application context, and (ii) not suitable, which means that the data construct is either unneeded or insufficiently represents the required information. Incomplete data constructs do not fulfil the latter requirement.

Table 5: Assessment of the suitability of the algorithms

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

4.3. Analysis of the empirical performance efficiency of the DCIP-MM

This analysis procedurally involved (i) specification of empirical performance indicators, (ii) generation of the empirical performance profiles of the prototype in the I-CPS applications, and (iii) exploring the performance limitations in the case of the three I-CPS applications. In terms of the performance indicators, the work extended to 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 were compared to the allowed computation time (ACT) in each of the application contexts, which is calculated by:

$$ACT = ART - (LTS + LTM) \quad (1)$$

Table 6: Assessment of the suitability of applied data constructs

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

Where: *ART* is the allowed respond time in a given application case, *LTS* and *LTM* are the practical latencies of information sensing and messaging, respectively. If the estimated computation time of an algorithm was higher than *ACT*, then the algorithm was considered as 'not effective'. The results of assessment are presented in Table 7.

Four factors that influence the practical performance of the systems in the target applications were considered, including (i) the accuracy of positioning, (ii) the latency of information sensing, (iii) length of personalized messages and (iv) the obedience of the informed user. Since these factors are strongly

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

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

depending on the applied techniques for practical application cases. The assessments were given according to the characteristics shown in Table 1&2. The investigation of the influential factors is summarized in Table 8.

4.4. Analysis of the theoretical performance sufficiency of the DCIP-MM

While the specific performance indicators could be captured quantitatively, due to their abstract and tentative nature, theoretical performance targets (TPTs) needed a qualitative evaluation. 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 totaled effort that is 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 overall performance targets could be achieved using the DCIP-MM, or what sort of limitations have been experienced. The second part of the investigations focused on the sustainability of the overall performance of the component computational mechanisms under the influence of various circumstances. The argumentation about the applicability of the DCIP-M 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 in the three target applications according to the following reasoning logic:

a. 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 is not fulfilled.
b. If either the testing of the processing algorithms, or the testing of the

Table 9: Principles for qualitative assessment of 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%

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

processed data constructs, or both, concludes with a negative result, then the proposed mechanism cannot be applied in the target application chases, since its empirical structural suitability is not validated.

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

d. 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.

On the basis of the results discussed above in Sections 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% with regard to 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:

$$\mathrm{OAI} = \sqrt[3]{\mathrm{TRI} * \mathrm{ESI} * \mathrm{EEI}} \qquad (2)$$

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 in the DCIP-MM is shown in Figure 6.

Three levels, including (i) high, (ii) medium, and (iii) low, are 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 high adaptation work before it can be applied in the target application. The second TPT (preparation effort) was assessed only based on the calculated ESI. Because, we believe that the preparation effort for using a computational mechanism in an application depends on how much change work with regard to the algorithms and data constructs is needed. Based on the propose principles, the assessment results of the TPTs for the target application cases are shown in Table 10.

5. Discussion, conclusions and future research

5.1. Reflections on the work and results

In the tradition of scientific inquiry, internal validation and external validation goes hand-in-hand. The VSA rests on the assumption that internal validity is guaranteed. 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. Focusing on usefulness of design methods, the VSA intended to filter out design methods, which could not meet the criterion of external validity, namely utility. Utility validation was driven by the question whether the method provided design solutions 'correctly' (effectiveness), and whether it provided 'correct' design solutions (efficiency). 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 obviously 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.

The above assumptions have been reused in the A-VSA, which was developed to validate the applicability of a computational mechanism, namely the DCIP-MM. Evidences about the measures of applicability were obtained with regards to its (i) functional/procedural relevance, (ii) suitability of the algorithm and the data constructs, (iii) the efficiency of the specific performances, and (iv) sustainability of the overall performance. If the DCIP-M fails in any one of these aspects, then the theoretical performance targets (sufficiency) cannot be fulfilled and there is no way to conclude positively about its applicability. If its relevance and suitability are positively evaluated, then the empirical performance plays a decisive role in the judgment. If the relevance and suitability indicators signalize a partial compliance and the efficiency designates only a partial completion, then the functional structure, the procedural flow, the algorithms and data structures, and the computational measures of the DCIP-MM should be adapted to the requirements of the specific application.

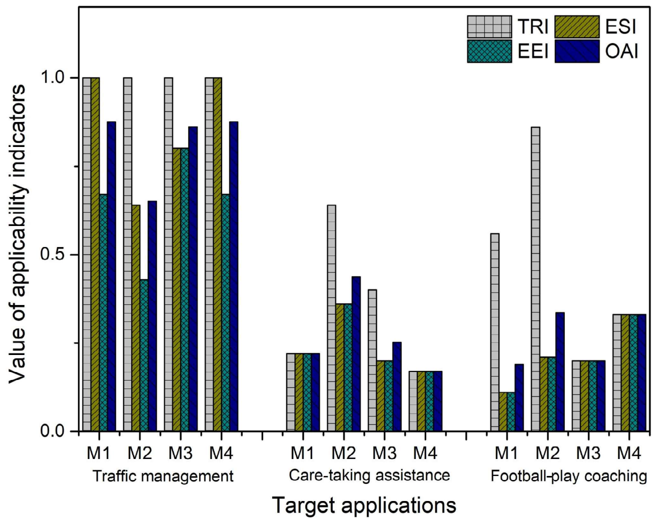

Fig. 6 Calculated indicators of the proposed mechanisms for the target applications

5.2. Conclusions

Over multiple applications in various studies, some deficiencies of the original VSA have been recognized, which are: (i) lack of explicit focus on formal verification and internal validation, (ii) relying dominantly on qualitative evaluation and argumentation concerning usefulness of a design method, (iii) focusing on the correctness and the utility of the results produced by design method, (iv) negligence of the role of information processing activities, the information/knowledge constructs, and designer competencies/experiences in judging the efficiency of design methods, and (v) limitation with regards to handling issues such as the influence of computational aspects on the usefulness of design methods. The A-VSA proposed in this paper is based on a combination of qualitative and quantitative perspectives, as well as of theoretical and empirical viewpoints. This makes it relevant for validation of complex computational mechanisms. For their application in cyber-physical systems, both structural appropriateness and computational performance need to be assessed from both theoretical and empirical points of view. In addition to using meaningful and reliable information, the A-VSA can also be used to explore the range (extent) of applicability.

Using the A-VSA, the applicability of a DCIP-MM has been assessed in terms of four aspects and various indicators related to them. The measure of theoretical structural appropriateness is relevance (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 (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 overall sufficiency, which is characterized by (i) the extent of achieving the overall performance targets, and (ii) the level of sustainable of the overall performance in varying circumstances.

On the one hand, the completed applicability-testing cross-case study confirmed that these measures are adequate and expressive enough. On the other hand, the results showed that the proposed DCIP-MM has a large application potential in each of the selected application cases. Notwithstanding, its appropriateness can be increased further by modifying some of its algorithms. However, some out-of-domain applications may raise the need for a more radical adaptation of the algorithms or even the functionality of the component mechanisms.

5.3. Future research opportunities

From the many opportunities of doing research in this particular field, we would like to expose three near-future research opportunities that are of importance:

- Combination of the A-VSA with rigorous verification of logical properness and validation of internal correctness
- Investigation if the concept of A-VSA can be used for external validation purposes other than usefulness and applicability validation
- Development of dedicated computational tools that support effective, systemic, and reliable application validation by software engineers.

Acknowledgments

The work is supported by “the Fundamental Research Funds for the Central Universities”.

References

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

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

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 Directions for Teaching and Learning 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. 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
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
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
Pedersen K, Emblemsvåg J, Bailey R, Allen JK, Mistree F (2000) Validating design methods & research: The validation square. In: 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
Sipser M (2006) Introduction to the theory of computation. Thomson Course Technology, Boston.
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
Schlesinger S (1979) Terminology for model credibility. Simul 32(3): 103-104. https://doi.org/10.1177/003754977903200304
Teegavarapu S (2019) Foundations of design method development. Dissertation, Clemson University.
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