

Exploiting Explicit Context Information for the Automatic Generation of Simulation Experiments

Dissertation

zur

Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

der Fakultät für Informatik und Elektrotechnik
der Universität Rostock



vorgelegt von

Pia Wilsdorf,
geb. am 19.08.1993
in Dresden

Rostock, 31. Juli 2024

Abstract

Modeling and simulation have been established as indispensable tool in science and engineering, offering means to explore and analyze complex systems, thereby enhancing the prediction and understanding of real-world phenomena. Developing valid models that accurately reproduce these phenomena is an intricate process, involving steps of model building intertwined with various simulation experiments for model calibration, validation, and analysis.

Especially conducting simulation experiments is challenging as their specification and execution vary depending, for example, on the diverse experiment types, methods, and tools required. This dissertation addresses the central question of how to support modelers in this intricate process, specifically, by automatically generating and executing simulation experiments. A pivotal strategy identified for achieving this goal is the generation of simulation experiments through reuse and adaption. The strategy can be subdivided into three parts, and the dissertation is organized accordingly.

The first part contributes to the explicit and automated specification of simulation experiments, employing model-driven engineering. In this approach, metamodels are crucial for formalizing the central ingredients of a simulation experiment and for defining a domain-specific language that allows representing simulation experiments in a tool-agnostic manner. This enables the automatic generation, transformation, and adaption of executable experiment specifications via a model-driven engineering pipeline.

The second part underlines the importance of context for understanding and conducting simulation studies, and particularly simulation experiments. A comprehensive and formalized definition of the conceptual model is introduced, which subsumes typical early-stage artifacts of a simulation study, from research objectives and assumptions to information sources, enriched with metadata. The conceptual model can be complemented by provenance, which is concerned with the process of generating and using the various artifacts. Formal approaches for representing provenance information are discussed that can illuminate the entire story of a simulation study, including the interrelations between artifacts, activities of model building and simulation experiments, and related studies. Open model databases are investigated as a key means for distributing, editing, and interconnecting provenance graphs and the conceptual model, ensuring their computational accessibility.

The third part focuses on utilizing the formal, machine-accessible representations to facilitate the automatic interpretation and exploitation of context information. This context information is integrated with the model-driven approach for experiment specification and execution to form the key components of the Reuse and Adapt framework for Simulation Experiments (RASE). In this framework, provenance patterns and inference rules enable the automatic initiation of an experiment generator depending on the last activities of a modeler, automatic identification of earlier simulation experiments to reuse, and their adaption to the context of a new simulation model. This supports the iterative nature of the modeling and simulation life cycle, enabling scenarios such as automatic regression testing and cross-validation of related models.

The results showcase how modelers can benefit from the developed approaches, emphasizing systematic, effective, and methodologically rigorous simulation studies while preserving user flexibility and control. Demonstrations in open-source software prototypes and diverse case studies, spanning stochastic discrete-event simulations, virtual prototyping, finite element analysis, and agent-based simulations, highlight the versatility and applicability of the proposed framework.

Finally, the dissertation identifies new research avenues towards fully automating simulation experiments as well as the entire modeling and simulation life cycle.

Zusammenfassung

Modellierung und Simulation haben sich als unverzichtbare Werkzeuge in Wissenschaft und Technik etabliert. Sie bieten Mittel zur Erforschung und Analyse komplexer Systeme und verbessern so die Vorhersage und das Verständnis von Phänomenen der realen Welt. Die Entwicklung valider Modelle, die diese Phänomene genau reproduzieren können, ist ein komplexer Prozess, der Schritte der Modellbildung umfasst, die mit verschiedenen Simulationsexperimenten zur Modellkalibrierung, -validierung und -analyse verflochten sind.

Insbesondere die Durchführung von Simulationsexperimenten stellt eine Herausforderung dar, da deren Spezifikation und Ausführung variiert, z. B. in Abhängigkeit von den verschiedenen Experimenttypen, Methoden und Werkzeugen, die benötigt werden. Diese Dissertation beschäftigt sich mit der zentralen Frage, wie Modelliererinnen und Modellierer in diesem komplexen Prozess unterstützt werden können, insbesondere durch die automatische Generierung und Ausführung von Simulationsexperimenten. Eine zentrale Strategie zur Erreichung dieses Ziels ist die Generierung von Simulationsexperimenten durch Wiederverwendung und Adaption. Die Strategie kann in drei Ansätze unterteilt werden, und die Dissertation ist entsprechend gegliedert.

Der erste Ansatz trägt zur expliziten und automatisierten Spezifikation von Simulationsexperimenten bei, indem er Methoden der modellgetriebenen Entwicklung einsetzt. In diesem Ansatz sind Metamodelle entscheidend für die Formalisierung der zentralen Bestandteile eines Simulationsexperiments und für die Definition einer domänenspezifischen Sprache, die es erlaubt, Simulationsexperimente in einer werkzeugunabhängigen Weise darzustellen. Dies ermöglicht die automatische Generierung, Transformation und Anpassung von ausführbaren Experimentspezifikationen über eine modellgetriebene Entwicklungspipeline.

Der zweite Ansatz unterstreicht die Bedeutung von Kontext für das Verständnis und die Durchführung von Simulationsstudien, insbesondere von Simulationsexperimenten. Es wird eine umfassende und formalisierte Definition des konzeptionellen Modells vorgestellt, das typische Artefakte der frühen Phase einer Simulationsstudie umfasst, von Forschungszielen und Annahmen bis hin zu Informationsquellen, angereichert mit Metadaten. Das konzeptionelle Modell kann durch Provenienz ergänzt werden, welche sich mit dem Prozess der Erzeugung und Verwendung der verschiedenen Artefakte befasst. Es werden formale Ansätze zur Darstellung von Provenienz-Informationen diskutiert, die die gesamte Geschichte einer Simulationsstudie repräsentieren können, einschließlich der Beziehungen zwischen Artefakten, Aktivitäten der Modellbildung und Simulationsexperimenten sowie verwandten Studien. Offene Modelldatenbanken werden als zentrales Mittel zur Verteilung, Bearbeitung und Vernetzung von Provenienz-Graphen und dem konzeptionellen Modell untersucht, um deren maschinelle Zugänglichkeit zu gewährleisten.

Der dritte Ansatz konzentriert sich auf die Nutzung der formalen, maschinell zugänglichen Darstellungen, um die automatische Interpretation und Verwertung von Kontextinformationen zu erleichtern. Diese Kontextinformationen werden mit dem modellgetriebenen Ansatz für die Spezifikation und Ausführung von Experimenten integriert und bilden die Schlüsselkomponenten des Reuse and Adapt Framework for Simulation Experiments (RASE). In diesem Rahmen ermöglichen Provenienz-Patterns und Inferenzregeln die automatische Initiierung eines Experimentgenerators in Abhängigkeit von den letzten Aktivitäten einer ModelliererIn oder eines Modellierers, die automatische Identifizierung früherer Simulationsexperimente zur Wiederverwendung und deren Anpassung an den Kontext eines neuen Simulationsmodells. Dies unterstützt die iterative Vorgehensweise im Modellierungs- und Simulationslebenszyklus und ermöglicht Szenarien wie automatische Regressionstests und Kreuzvalidierung verwandter Modelle.

Die Ergebnisse zeigen, wie Modelliererinnen und Modellierer von den entwickelten Ansätzen

profitieren können, wobei der Schwerpunkt auf systematischen, effektiven und methodisch präzisen Simulationsstudien liegt und gleichzeitig die Flexibilität und Kontrolle der Benutzerinnen und Benutzer erhalten bleibt. Demonstrationen in Open-Source-Software-Prototypen und verschiedenen Fallstudien, welche stochastische diskret-ereignisorientierte Simulationen, virtuelles Prototyping, Finite-Elemente-Analyse und agentenbasierte Simulationen umfassen, unterstreichen die Vielseitigkeit und Anwendbarkeit des vorgestellten Frameworks.

Abschließend identifiziert die Dissertation neue Forschungsrichtungen zur vollständigen Automatisierung von Simulationsexperimenten sowie des gesamten Modellierungs- und Simulationslebenszyklus.