

Alexander Detzner

**Digital Twin and Data Analytics
for Product Quality Monitoring
and Root-Cause Analysis**

Schriftenreihe VPE

Band 29

Herausgeber: Prof. Dr.-Ing. Martin Eigner

Digital Twin and Data Analytics for Product Quality Monitoring and Root-Cause Analysis

vom Fachbereich Maschinenbau und Verfahrenstechnik
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

genehmigte

Dissertation

von

Herrn

M.Sc. Alexander Clemens Detzner

aus Goslar

Dekan: Prof. Dr.-Ing. Tilmann Beck
Vorsitzender der Prüfungskommission: Prof. Dr.-Ing. Jens C. Göbel
1. Berichterstatter: Prof. Dr.-Ing. Martin Eigner
2. Berichterstatter: Prof. Dr.-Ing. Oliver Riedel
Tag der mündlichen Prüfung: 29.03.2022

Kaiserslautern 2022

D 386

Alexander Detzner

**Digital Twin and Data Analytics for Product Quality
Monitoring and Root-Cause Analysis**

D 386 (Diss. Technische Universität Kaiserslautern)

Shaker Verlag
Düren 2022

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: Kaiserslautern, TU, Diss., 2022

Copyright Shaker Verlag 2022

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-8719-2

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren

Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9

Internet: www.shaker.de • e-mail: info@shaker.de

Kurzfassung

Die Komplexität des Produktentwicklungsprozesses ist in den letzten Jahren kontinuierlich angestiegen. Neben den immer komplexeren Lieferketten zwischen Herstellern und deren Lieferanten, verteilten Produktionsstandorten und einer zunehmenden Individualisierung der Produkte, stellt vor allem die Entwicklung zu an das Internet angebotenen, cyber-physischen Produkten und dazugehörigen Services eine fundamentale Neuerung dar.

Neben den Herausforderungen für die Produktentwickler, ist die gestiegene Komplexität insbesondere im Qualitätsmanagement bemerkbar. Hochvernetzte Systeme und Prozesse resultieren in Problemen, die sich mit präventiven Maßnahmen nicht vollumfänglich abfangen lassen. Aus diesem Grund können Probleme beim Kunden nicht komplett verhindert werden. Im Falle eines solchen Problems, sind jedoch derart viele Informationen und Daten verfügbar, dass sich die Suche nach der technischen Fehlerursache, der Root-Cause, oftmals zeitaufwendig und schwierig gestaltet.

In diesem Spannungsfeld benötigen Produktentwickler und Qualitätsingenieure sowohl ein effizientes Management der verfügbaren Daten über jede einzelne Produktinstanz, als auch Software, welche die Root-Cause Analyse durch (teilweise) Automatisierung unterstützt und vereinfacht. Diese Dissertation diskutiert daher einen Digitalen Zwilling für das Qualitätsmanagement, sowie Datenanalyse Methoden zur Root-Cause Analyse.

Basierend auf einer detaillierten Literaturstudie, wird eine ganzheitliche Definition für den Digitalen Zwilling entwickelt und anschließend auf den konkreten Anwendungsfall des Qualitätsmanagements in der Automobilbranche zugeschnitten. Dieser Digitale Zwilling besteht, neben einem strukturierenden Digital Twin Core, aus Supplemental Data, Physical Twin Data, Process Data, Service Data und Generated Data.

Im Anschluss werden verschiedene Mustererkennungsmethoden aus dem Bereich der künstlichen Intelligenz verglichen, um die ideale Methode zur Anwendung in der Root-Cause Analyse zu bestimmen. Dabei wird der sog. Information Gain als bevorzugte Methode identifiziert. Zuletzt fließen die Erkenntnisse in die Entwicklung einer Software zur Root-Cause Analyse ein, in der mithilfe eines interaktiven Entscheidungsbaumes die Ursachen und Zusammenhänge im Kontext eines Produktdefektes visualisiert werden können.

Abstract

The complexity of the product development process has been increasing continuously over the last decades. The reasons are progressively complex supply chains as well as distributed production sites and an increasing individualization of products. Furthermore, the development of cybertronic products connected to the internet as well as the software and services associated with these products present new challenges and require fundamental innovations in product development.

In addition to the challenges for product developers, the increased complexity is relevant for quality management. Highly connected systems and processes result in problems that cannot be fully controlled with preventive measures. For this reason, product failures cannot be completely avoided. In case of a problem, such vast amounts of data are available that the search for the technical cause of the error - the root-cause - is often time-consuming and difficult.

Given these challenges, product developers and quality engineers require an efficient management of the available data about each individual product instance and software that supports their root-cause analysis through (partial) automation. Therefore, this dissertation presents a Digital Twin for quality management, as well as data analysis methods for root-cause analysis.

Based on a detailed literature study, a holistic definition for the Digital Twin is developed and then detailed for the use case of quality management in the automotive domain. This presented Digital Twin consists of a structuring Digital Twin Core as well as Supplemental Data, Physical Twin Data, Process Data, Service Data and Generated Data.

Subsequently, different pattern recognition methods from the field of artificial intelligence are compared to determine the ideal method for application in root-cause analysis. The Information Gain is identified as the preferable method. Finally, the findings are incorporated into the development of a software for root-cause analysis, which visualizes correlations and relationships regarding a product defect with the help of an interactive decision tree.

Table of Contents

Kurzfassung	i
Abstract	ii
Table of Contents	iii
Abbreviations	vii
Figures	ix
Tables	x
1 Introduction	1
1.1 Motivation	2
1.2 Research Goals.....	5
1.3 Overview	6
2 State of the Art	9
2.1 Reactive Quality Management	9
2.1.1 Product Monitoring in the Automotive Domain.....	9
2.1.2 Field Data Analytics	13
2.2 Product Lifecycle Management	15
2.2.1 The Product Lifecycle	16
2.2.2 PLM Definition	17
2.2.3 PLM Architecture	18
2.2.4 Variant Management	21
2.3 Knowledge Discovery in Databases and Machine Learning.....	23
2.3.1 Terms and Definitions	24
2.3.2 Machine Learning Approaches	26
2.4 Pattern Recognition in Product Lifecycle Management	27
2.4.1 Non-destructive Classification of Product or Process Quality	27
2.4.2 Predictive Maintenance of Products in the Field	28
2.4.3 Root-cause Analysis and Diagnostics in Reactive Quality Management	29
2.4.4 Related Research in Pattern Recognition	33

3	The Digital Twin Concept	35
3.1	Short History of the Digital Twin	36
3.2	Discussion about the Digital Twin Concept	38
3.2.1	Derivation of a Definition	38
3.2.2	Digital Twin Core	41
3.2.3	Supplemental Data	42
3.2.4	Process Data	42
3.2.5	Service Data	43
3.2.6	Physical Twin Data	44
3.2.7	Generated Data	44
3.3	The Digital Twin throughout the Product Lifecycle	45
3.3.1	Lifecycle of a Digital Twin	45
3.3.2	Digital Thread	47
4	A Digital Twin for Automotive Quality Monitoring.....	49
4.1	Elements of the Digital Twin	50
4.1.1	Product Configuration	50
4.1.2	Part Traceability	52
4.1.3	Process Monitoring	54
4.1.4	Service Reports	56
4.1.5	Usage Data	57
4.2	Architecture	61
4.2.1	Integrating Data Sources	61
4.2.2	Hierarchical Data Structure	64
5	Root-Cause Analysis with Feature Selection Methods.....	71
5.1	Methods for feature selection	72
5.1.1	Filter approaches	73
5.1.2	Embedded approaches	75
5.1.3	Wrapper approaches	76
5.1.4	Other approaches	76
5.2	Properties of top-level product attributes	78
5.2.1	Imbalanced classes	79
5.2.2	High-dimensional feature space	80

5.2.3	Inhomogeneous classes	81
5.2.4	Global and local correlations	82
5.2.5	Non-deterministic failure	83
5.3	Simulation Study	85
5.3.1	Generating Features	85
5.3.2	Assigning Class Labels	87
5.4	Evaluation of feature ranking methods	89
5.4.1	Requirements for root-cause analysis	89
5.4.2	Comparison of feature ranking methods	93
5.5	Conclusion	100
6	Interactive Root-Cause Analysis.....	101
6.1	Interactive Decision Trees	101
6.1.1	Preceding Research on Interactivity in Decision Trees.....	101
6.1.2	Interactive Decision Trees for Root-Cause Analysis.....	102
6.2	Software Prototype Implementation	104
6.2.1	Design Process.....	104
6.2.2	Choice of Reference	105
6.2.3	User Interface	106
6.2.4	Control Options.....	109
7	Conclusion.....	113
7.1	Summary	113
7.2	Outlook.....	115
	Student research projects that contributed to the dissertation.	117
	Literature	119
	Curriculum Vitae	133

Abbreviations

AI	Artificial Intelligence
Als	Affected Items
BMW	Bayerische Motoren Werke
BOM	Bill of Materials
BOP	Bill of Processes
CAD	Computer Aided Design
CAM	Computer Aided Manufacturing
CAE	Computer Aided Engineering
CASE	Computer Aided Software Engineering
CI	Configuration Items
CIM	Computer Integrated Manufacturing
CM	Configuration Management
CMIM	Conditional Mutual Information Maximization criterion
CPU	Central Processing Unit
CRM	Customer Relationship Management
DTC	Diagnostic Trouble Code
ECM	Engineering Change Management
ECU	Electronic Control Unit
EDM	Engineering Data Management
EoL	End of Line
ERP	Enterprise Resource Planning
FE	Finite Element
FMEA	Failure Mode and Effects Analysis
GLM	Generalized Linear Models
GPS	Global Positioning System

ISO	International Standards Organization
IoT	Internet of Things
IT	Information Technology
KDD	Knowledge Discovery in Databases
MBSE	Model Based Systems Engineering
MES	Manufacturing Execution System
ML	Machine Learning
MRO	Maintenance Repair and Overhaul
OEM	Original Equipment Manufacturer
OLAP	Online Analytical Processing
PCA	Principal Component Analysis
PDM	Product Data Management
PLM	Product Lifecycle Management
PPS	Production Planning and Steering
QM	Quality Management
RFID	Radio Frequency IDentification
RM	Requirements Management
SCM	Supply Chain Management
SOC	State of Charge
STEP	STandard for the Exchange of Product model data
TDM	Team Data Management
UID	Unique Identifier Number
US	United States (of America)
UX	User eXperience
VIN	Vehicle Identifier Number

Figures

Figure 1.1 Feedback of lessons learned during the usage phase	2
Figure 1.2 Examples for Product Lifecycle Data Silos	3
Figure 1.3 A Digital Twin for Product Lifecycle Data.....	4
Figure 2.1 Defect code of warranty claim in the automotive domain	10
Figure 2.2 Example of a quality trend line	11
Figure 2.3 Warranty claim data.....	12
Figure 2.4 Quality information feedback loop	13
Figure 2.5 Multidimensional data cube [Gab11]	14
Figure 2.6 Multi-dimensional product development process [Eig09].....	15
Figure 2.7 Product lifecycle [Vaj14]	16
Figure 2.8 Data sources throughout the product lifecycle [Kas15].....	17
Figure 2.9 IT-architecture for enterprise PLM-solutions [Eig21a].....	20
Figure 2.10 The Digital Thread across product lifecycle phases [Eig21a]	21
Figure 2.11 Generic Product Structure, adapted from [Eig09].....	22
Figure 3.1 Generalized Model of a Digital Twin, amended based on [Eig19; Eig21b] ..	41
Figure 3.2 Digital Twin throughout the product lifecycle [Eig21a]	47
Figure 3.3 The Digital Thread connecting Digital Twin and Digital Model [Eig21a].....	48
Figure 4.1 Anomaly detection based on Digital Twin data	49
Figure 4.2 As-designed BOM for the Digital Twin.....	51
Figure 4.3 Basic Digital Twin Core	51
Figure 4.4 Digital Twin based on the as-built BOM including Supplemental Data	53
Figure 4.5 Enrichment of Digital Twin along manufacturing and logistics processes...	54
Figure 4.6 Bill of materials per part with corresponding process data.	55
Figure 4.7 Enrichment of Digital Twin with maintenance, repair and overhaul data ...	57
Figure 4.8 Data storage within the Enterprise IoT Platform	59
Figure 4.9 Aggregation of usage phase data in the vehicle.....	60
Figure 4.10 Elements of the Digital Twin	61
Figure 4.11 Tracing information throughout the entire product lifecycle.	64
Figure 4.12 Aggregation and refinement of data in a Digital Twin	67
Figure 5.1 Graph representation of node neighborhood [Pea09].....	77
Figure 5.2 Types and properties of top-level product attributes.....	79
Figure 5.3 Spurious correlation between random variables [Vig17].....	80
Figure 5.4 Root-causes and influences of a perceived problem	82

Figure 5.5 Simulating product attributes	87
Figure 5.6 Ranking positions of filter methods.....	94
Figure 5.7 Ranking positions of features with correlations	96
Figure 5.8 Ranking positions of features with correlations	98
Figure 6.1 Decision tree in Interactive Decision Tree Software [Blu10]	103
Figure 6.2 Attribute table in interactive decision tree software [Blu10]	104
Figure 6.3 Definition of defective and reference vehicles	106
Figure 6.4 Node representation in interactive decision tree.....	107
Figure 6.5 Attribute frequency in defective vehicles and reference	108
Figure 6.6 Table of product attributes in interactive decision tree.....	109
Figure 6.7 Tree with drag-and-drop and multi-select option	112

Tables

Table 2.1 Feature-Instance-Matrix.....	24
Table 5.1 Example for Simpson's paradox	83
Table 5.2 Class labels without feature interaction	88
Table 5.3 Class labels including feature interaction.....	89
Table 5.4 Frequency of occurrence in Markov Blanket.....	99