

Berechnung und Optimierung von Grenzwerten für Mengen von Software Metriken

Steffen Herbold, Jens Grabowski, Stephan Waack
Georg-August-Universität Göttingen



Softwareforen Leipzig

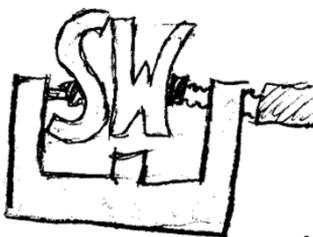
Womit wir uns beschäftigen ...

Testsprachen

- TTCN-3
- UML Testprofil

Qualitätssicherung (ohne Testen) für

- TTCN-3 Testspecs
- UML



Testmethoden

- IOP Testen von Grids und Clouds
- Testauswahl basierend auf Benutzungsprofilen

Lernverfahren in der Qualitätssicherung

- Bewertung von Metriken
- Systemrekonstruktion aus Traces (Tests)

Überblick

- ▶ Motivation
- ▶ Software Metriken
- ▶ Klassifikation durch Grenzwerte
- ▶ Optimierung von Metrik Mengen
- ▶ Anwendungen
- ▶ Fallstudien
- ▶ Zusammenfassung

▶ 3

Optimierung von Metrik Mengen mit Grenzwerten

Was ist Software Engineering

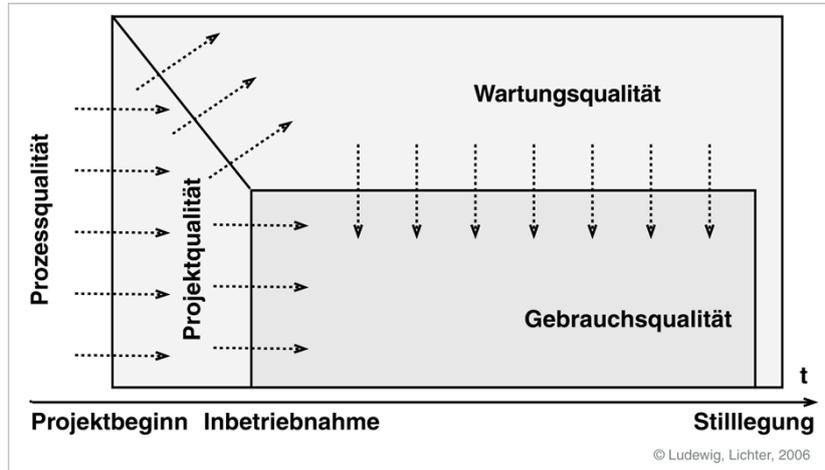
- ▶ Software Engineering ist:
 - ▶ die Entwicklung
 - ▶ die Pflege und
 - ▶ der Einsatz
- qualitativ hochwertiger Software unter Einsatz von
 - ▶ wissenschaftlichen Methoden
 - ▶ wirtschaftlichen Prinzipien
 - ▶ geplanten Vorgehensmodellen
 - ▶ Werkzeugen und
 - ▶ quantifizierbaren Zielen.

□ B. Kahlbrandt: *Software-Engineering: Objektorientierte Software-Entwicklung mit der Unified Modeling Language*, Springer Verlag (1998)

▶ 4

Optimierung von Metrik Mengen mit Grenzwerten

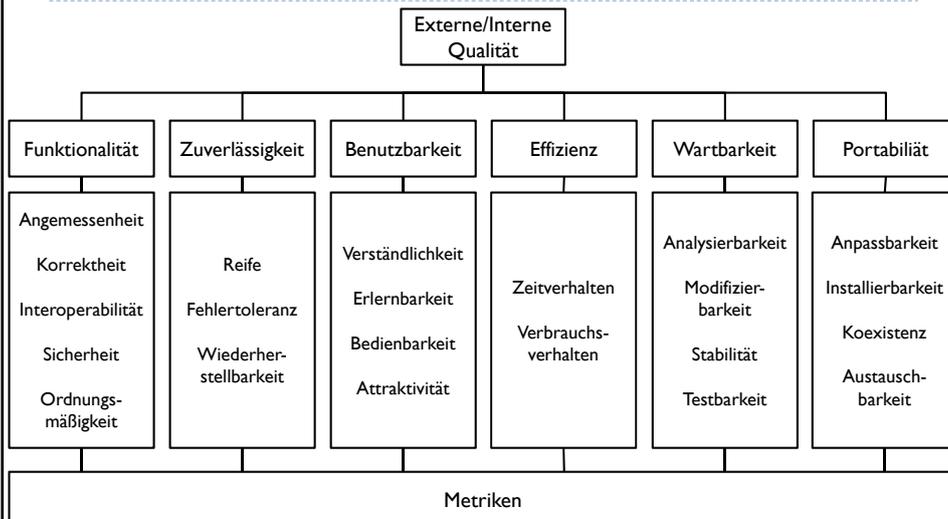
Zusammenhang zwischen Arten von Qualität



▶ 5

Optimierung von Metrik Mengen mit Grenzwerten

Qualitätsbewertung nach ISO 9126



▶ 6

Optimierung von Metrik Mengen mit Grenzwerten

Überblick

- ▶ Motivation
- ▶ **Software Metriken**
- ▶ Klassifikation durch Grenzwerte
- ▶ Optimierung von Metrik Mengen
- ▶ Anwendungen
- ▶ Fallstudien
- ▶ Zusammenfassung

▶ 7

Optimierung von Metrik Mengen mit Grenzwerten

Software Metriken

- ▶ “You cannot control what you cannot measure” (Tom DeMarco)
- ▶ “To measure is to know” (Clerk Maxwell)



Leistung	100 PS
Verbrauch	5,8 l
Geschwindigkeit	176 km/h
Gewicht	1458 kg



???

▶ 8

Optimierung von Metrik Mengen mit Grenzwerten

Eigenschaften von Metriken

- ▶ Messungsarten
 - ▶ Intern
 - ▶ Extern

- ▶ Messobjekte
 - ▶ Produkte
 - ▶ Prozesse
 - ▶ Ressourcen

- ▶ Statische Quelltextanalyse

▶ 9

Optimierung von Metrik Mengen mit Grenzwerten

Metriken für Methoden und Klassen

Module, Dateien

Klassen

Coupling Between Objects (CBO)
 Response For a Class (RFC)
 Weighted Method per Class (WMC)
 Number of Overriden Methods (NORM)
 Lines Of Code (LOC)
 Number Of Methods (NOM)
 Number of Static Methods (NSM)

Methoden

Number of Statements (NST)
 McCabe's Cyclomatic Number (VG)
 Nested Block Depth (NBD)
 Number of Function Calls (NFC)

▶ 10

Optimierung von Metrik Mengen mit Grenzwerten

Metriken für Methoden

- ▶ **Number of Statements (NST)**
 - ▶ Anzahl der Anweisungen im Quelltext
- ▶ **McCabe's Cyclomatic Number (VG)**
 - ▶ Anzahl der Verzweigungen im Kontrollfluss
- ▶ **Nested Block Depth (NBD)**
 - ▶ Verschachtelungstiefe
- ▶ **Number of Function Calls (NFC)**
 - ▶ Anzahl der Methoden die Aufgerufen werden

▶ 11

Optimierung von Metrik Mengen mit Grenzwerten

Metriken für Klassen (I)

- ▶ **Coupling Between Objects (CBO)**
 - ▶ Anzahl der Klassen, mit denen eine Klasse zusammenhängt
- ▶ **Response For a Class (RFC)**
 - ▶ Anzahl der Methoden, die Aufgerufen werden, wenn die Methoden einer Klasse aufgerufen werden
- ▶ **Weighted Methods per Class (WMC)**
 - ▶ Summe der Komplexität der Methoden
 - ▶ McCabe's Cyclomatic Number (VG)

▶ 12

Optimierung von Metrik Mengen mit Grenzwerten

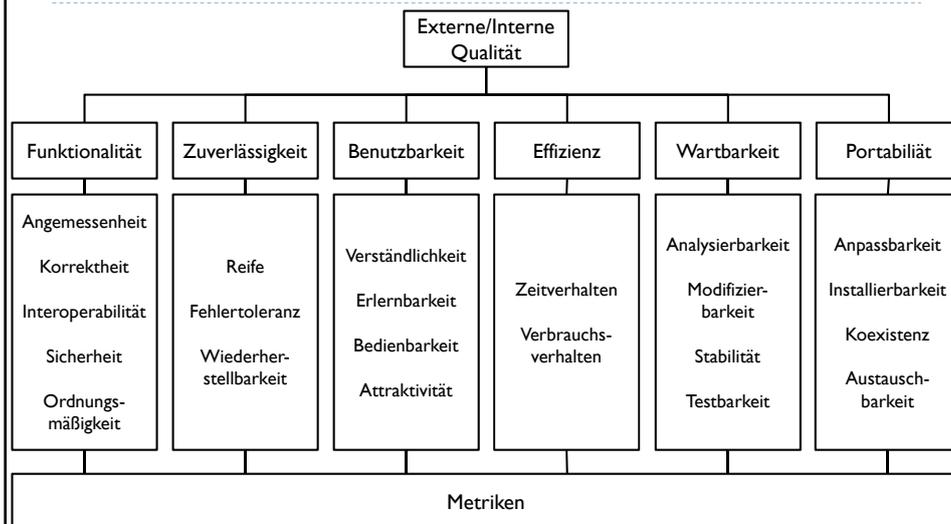
Metriken für Klassen (II)

- ▶ **Number of Overriden Methods (NORM)**
 - ▶ Anzahl der Methoden aus einer Oberklasse, die überschrieben wurden
- ▶ **Lines of Code (LOC)**
 - ▶ Anzahl der Quelltextzeilen (ohne Leerzeilen und Kommentare)
- ▶ **Number of Methods (NOM)**
 - ▶ Anzahl der Methoden
- ▶ **Number of Static Methods (NSM)**
 - ▶ Anzahl der statischen Methoden

▶ 13

Optimierung von Metrik Mengen mit Grenzwerten

Qualitätsbewertung nach ISO 9126



▶ 14

Optimierung von Metrik Mengen mit Grenzwerten

Überblick

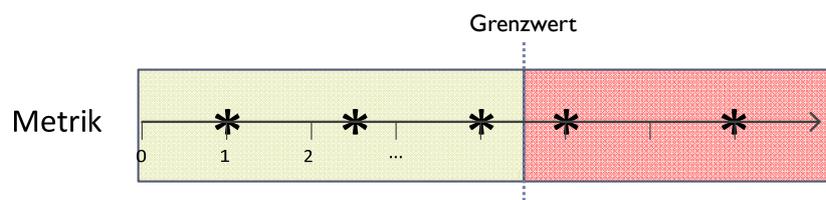
- ▶ Motivation
- ▶ Software Metriken
- ▶ **Klassifikation durch Grenzwerte**
- ▶ Optimierung von Metrik Mengen
- ▶ Anwendungen
- ▶ Fallstudien
- ▶ Zusammenfassung

▶ 15

Optimierung von Metrik Mengen mit Grenzwerten

Grenzwerte

- ▶ Methode zum Trennen von Werten



- ▶ Obere oder untere Grenze
 - ▶ Betrachten nur obere Grenzen

▶ 16

Optimierung von Metrik Mengen mit Grenzwerten

Grenzwerte für Methoden

Metrikname	Programmiersprache	Grenzwert
McCabe's Cyclomatic Number (VG)	C	24
	C++/C#	10
Nested Block Depth (NBD)	C/C++/C#	5
Number of Function Calls (NFC)	C/C++/C#	5
Number of Statements (NST)	C/C++/C#	50

▶ 17

Optimierung von Metrik Mengen mit Grenzwerten

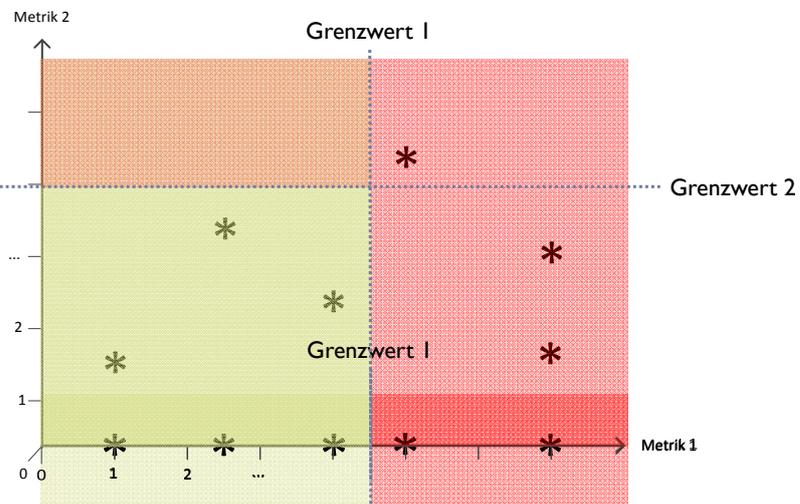
Grenzwerte für Java Klassen

Metrikname	Grenzwert
Weighted Methods per Class (WMC)	100
Coupling Between Objects (CBO)	5
Response For a Class (RFC)	100
Number of Overriden Methods (NORM)	3
Lines of Code (LOC)	500
Number of Methods (NOM)	20
Number of Static Methods (NSM)	4

▶ 18

Optimierung von Metrik Mengen mit Grenzwerten

Grenzwerte und Rechtecke



▶ 19

Optimierung von Metrik Mengen mit Grenzwerten

Überblick

- ▶ Motivation
- ▶ Software Metriken
- ▶ Klassifikation durch Grenzwerte
- ▶ **Optimierung von Metrik Mengen**
- ▶ Anwendungen
- ▶ Fallstudien
- ▶ Zusammenfassung

▶ 20

Optimierung von Metrik Mengen mit Grenzwerten

Grundidee

- ▶ Rechtecke = Grenzwertemengen
- ▶ Rechteck durch maschinelles Lernen bestimmen
- ▶ Datenbasiertes Verfahren
 - ▶ Bereits gemessene Software mit Aufteilung in Gut/Schlecht

▶ 21

Optimierung von Metrik Mengen mit Grenzwerten

Optimierung von Metrik Mengen

- ▶ Gegeben
 - ▶ Metrik Menge $M = \{m_1, \dots, m_n\}$
 - ▶ Softwaremodule $S = \{s_1, s_2, \dots\}$
 - ▶ Metrikwerte $m_1(s_i), \dots, m_n(s_i)$
 - ▶ Klassifikation $f(s_i)$
- ▶ Gesucht
 - ▶ Teilmenge von $M^* \subseteq M$ inkl. Grenzwerten, so dass
 - ▶ $f_{M^*}(s_i) \approx f(s_i)$
 - ▶ $|M^*|$ minimal

▶ 22

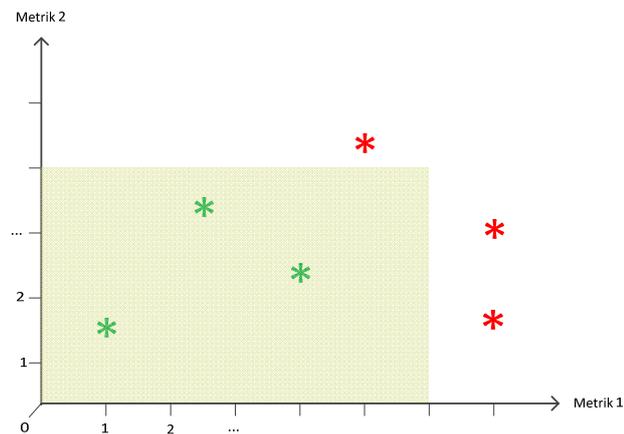
Optimierung von Metrik Mengen mit Grenzwerten

Grenzwertberechnung

- ▶ Berechne für alle Teilmengen Grenzwerte

- ▶ $\{m_1\}, \{m_1, m_2\}, \{m_1, m_3\}, \dots, \{m_1, \dots, m_n\}$

- ▶ 2^n viele



▶ 23

Optimierung von Metrik Mengen mit Grenzwerten

Auswahl der besten Teilmenge

- ▶ Klassifikationsfehler ϵ bestimmen

- ▶ Abweichung von der Teilmenge zur Eingabe
 - ▶ Wahrscheinlichkeit das $f_{M^*}(s_i) \neq f(s_i)$

- ▶ Wähle kleinste Teilmenge mit hinreichend kleinem ϵ

- ▶ $\epsilon \leq \delta$ für ein gesetztes Fehlerlimit δ
 - ▶ $\delta = 1\%$
 - ▶ δ in 0.5% erhöhen, bis eine Teilmenge gefunden wird

▶ 24

Optimierung von Metrik Mengen mit Grenzwerten

Überblick

- ▶ Motivation
- ▶ Software Metriken
- ▶ Klassifikation durch Grenzwerte
- ▶ Optimierung von Metrik Mengen
- ▶ **Anwendungen**
- ▶ Fallstudien
- ▶ Zusammenfassung

▶ 25

Optimierung von Metrik Mengen mit Grenzwerten

Anwendungen

- ▶ Reduzierung der Größe von Metrik Mengen
 - ▶ Höhere Effizienz
- ▶ Grenzwerte statt anderem Klassifikator einführen
 - ▶ Bessere Interpretierbarkeit
- ▶ Einführen von Umgebungsspezifischen Grenzwerten
 - ▶ Automatisierbare Qualitätsanalyse

▶ 26

Optimierung von Metrik Mengen mit Grenzwerten

Größenreduktion

- ▶ Metrik Menge M mit Grenzwerten vorhanden
- ▶ Klassifiziere die Software S mit Hilfe von M
- ▶ Berechne optimale Teilmenge $M^* \subseteq M$
 - ▶ M^* effizienter als M

▶ 27

Optimierung von Metrik Mengen mit Grenzwerten

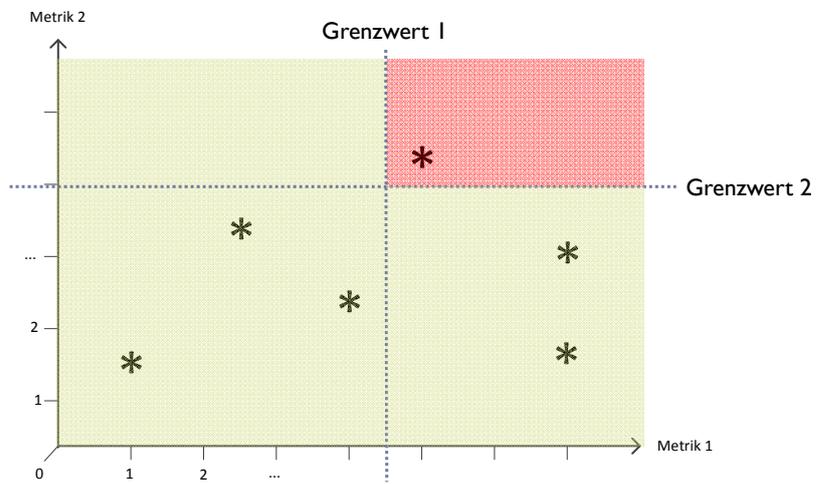
Vereinfachung der Klassifikation

- ▶ Grenzwerte statt anderem Klassifikator verwenden
 - ▶ Verstöße gegen Grenzwerte zugelassen
 - ▶ Entscheidungsbäume
- ▶ Klassifiziere Software S durch Klassifikator $f_{complex}$
- ▶ Wähle geeignete Kandidaten Metriken M
- ▶ Berechne optimale Teilmenge $M^* \subseteq M$

▶ 28

Optimierung von Metrik Mengen mit Grenzwerten

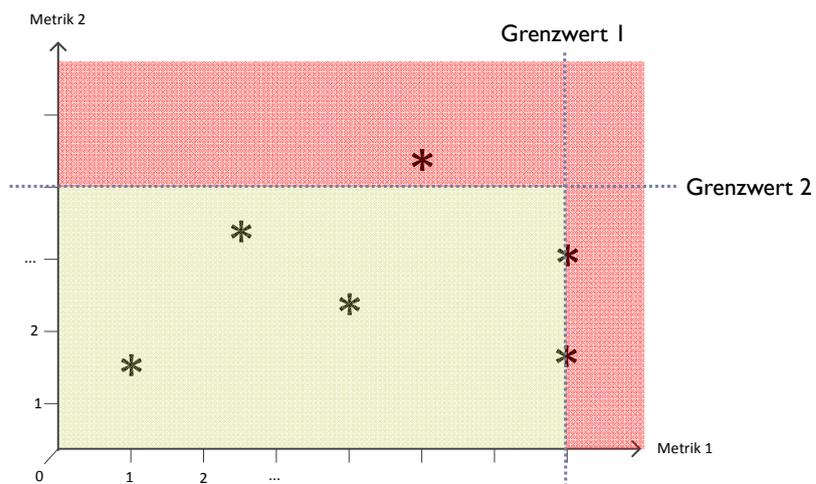
Klassifikation mit einem erlaubtem Verstoß



▶ 29

Optimierung von Metrik Mengen mit Grenzwerten

Approximation durch Grenzwerte



▶ 30

Optimierung von Metrik Mengen mit Grenzwerten

Umgebungsspezifische Grenzwerte

- ▶ **Annahme**
 - ▶ Noch kein Klassifikator vorhanden

- ▶ **Experte liefert Grundlagen**
 - ▶ Manuell Klassifikation einer Teilmenge der Software S
 - ▶ Auswahl von Kandidatenmetriken M

- ▶ **Berechne optimale Teilmenge $M^* \subseteq M$**

▶ 31

Optimierung von Metrik Mengen mit Grenzwerten

Überblick

- ▶ Motivation
- ▶ Software Metriken
- ▶ Klassifikation durch Grenzwerte
- ▶ Optimierung von Metrik Mengen
- ▶ Anwendungen
- ▶ **Fallstudien**
- ▶ Zusammenfassung

▶ 32

Optimierung von Metrik Mengen mit Grenzwerten

Datengrundlage

► Auf Basis von 8 Open Source Projekten

Name	Version	Programmiersprache	Größe
Apache Webserver	2.2.10	C	6718 Methoden
kdebase	12/05/2008	C++	21404 Methoden
kdelibs	12/05/2008	C++	37444 Methoden
AspectDNG	1.0.3	C#	2759 Methoden
NetTopologySuite	1.7.1.RC1	C#	3059 Methoden
SharpDevelop	2.2.1.2648	C#	15700 Methoden
Eclipse Java Development Tools	3.2	Java	4833 Klassen
Eclipse Platform Project	3.2	Java	5399 Klassen

► 33

Optimierung von Metrik Mengen mit Grenzwerten

Fallstudie: Optimierung von Metrik Mengen

► C Funktionen

	VG	NBD	NFC	NST
Eingang	24	5	5	50
Optimiert			5	

0.78% Fehler

75% Größenreduktion!

► C++ Methoden und C# Methoden

	VG	NBD	NFC	NST
Eingang	10	5	5	50
Optimiert			5	

0.06% Fehler, C++

0.59% Fehler, C#

75% Größenreduktion!

► Java Klassen

	WMC	CBO	RFC	NORM	LOC	NOM	NSM
Eingang	100	5	100	3	500	20	4
Optimiert		5		3			4

0.27% Fehler

57% Größenreduktion!

► 34

Optimierung von Metrik Mengen mit Grenzwerten

Verwenden eines anderen Klassifikators (1)

▶ C Funktionen – 1 Verstoß erlaubt

	VG	NBD	NFC	NST
Eingang	24	5	5	50
Optimiert				50

0.84% Fehler

▶ C++ Methode – 1 Verstoß erlaubt

	VG	NBD	NFC	NST
Eingang	10	5	5	50
Optimiert	10			

75% Größenreduktion!

0.87% Fehler

▶ C# Methoden – 1 Verstoß erlaubt

	VG	NBD	NFC	NST
Eingang	10	5	5	50
Optimiert	9			

1.26% Fehler

▶ 35

Optimierung von Metrik Mengen mit Grenzwerten

Verwenden eines anderen Klassifikators (2)

▶ Java Klassen – 1 Verstoß erlaubt

	WMC	CBO	RFC	NORM	LOC	NOM	NSM
Eingang	100	5	100	3	500	20	4
Optimiert			98	3		20	4

1.71% Fehler

42% Größenreduktion!

▶ Java Klassen – 2 Verstöße erlaubt

	WMC	CBO	RFC	NORM	LOC	NOM	NSM
Eingang	100	5	100	3	500	20	4
Optimiert	99		110				

2.21% Fehler

71% Größenreduktion!

▶ 36

Optimierung von Metrik Mengen mit Grenzwerten

Fallstudienresultate

- ▶ Größe von Metrik Mengen erfolgreich optimiert
 - ▶ 42%-75% kleinere Mengen

- ▶ Fehler im Bereich des Rauschens
 - ▶ Statistische Tests haben auch keine Schwächen gezeigt

- ▶ Komplexere Klassifikation durch Grenzwerte ersetzt

▶ 37

Optimierung von Metrik Mengen mit Grenzwerten

Überblick

- ▶ Motivation
- ▶ Software Metriken
- ▶ Klassifikation durch Grenzwerte
- ▶ Optimierung von Metrik Mengen
- ▶ Anwendungen
- ▶ Fallstudien
- ▶ **Zusammenfassung**

▶ 38

Optimierung von Metrik Mengen mit Grenzwerten

Zusammenfassung

- ▶ **Metrik Mengen mit Grenzwerten zur Qualitätsbewertung**
 - ▶ Simple Verfahren mit hoher Effektivität
- ▶ **Datenbasiertes bestimmen von Grenzwerten**
 - ▶ Basierend auf maschinellen Lernen
- ▶ **Komplexere Verfahren durch Grenzwerte zu Ersetzbar**
 - ▶ Bessere Interpretierbarkeit
- ▶ **Fallstudien belegen, dass i.d.R. wenige Metriken ausreichen**
 - ▶ Weniger Aufwand für die Datensammlung

▶ 39

Optimierung von Metrik Mengen mit Grenzwerten

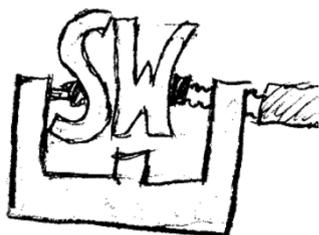
Ausblick

- ▶ **Disjunktive Normalformen statt einfachen Grenzwerten**
 - ▶ $(m_1 \wedge m_2) \vee (m_1 \wedge m_3 \wedge m_4)$
- ▶ **Rating statt Klassifikation**
 - ▶ Kritisch, Suspekt, Unproblematisch
- ▶ **Metrik Mengen auf anderen Abstraktionsebenen**
 - ▶ Module, Projekte
- ▶ **Einbeziehung von Prozess- und Ressourcenmetriken**
 - ▶ Fehlerzahl, Testaufwand

▶ 40

Optimierung von Metrik Mengen mit Grenzwerten

Vielen Dank für ihre Aufmerksamkeit



Jens Grabowski
grabowski@informatik.uni-goettingen.de

Steffen Herbold
herbold@informatik.uni-goettingen.de