Evaluation adaptiver Entwurfsmangelerkennung

Jochen Kreimer, jotte@uni-paderborn.de
Universität Paderborn, Institut für Informatik
Fürstenallee 11, 33102 Paderborn

1 Einleitung

Die Qualität von Software kann je nach Anwendungsgebiet an unterschiedlichen Kriterien gemessen werden. Für große Software-Systeme spielen u. a. Kriterien wie Wartbarkeit, Verständlichkeit und Erweiterbarkeit eine wichtige Rolle.

Mein Ziel ist es, Entwurfsmängel in Software-Systemen zu erkennen und somit „schlechte“ — unverständliche, schwer erweiter- und änderbare — Programmstrukturen zu vermeiden. Prominente Entwurfsmängel sind z. B. die von Fowler eingeführten *Bad Smells* in objektorientierten Programmen.

Entwurfsmängel werden abhängig von der Sichtweise und dem Erfahrungsschatz des Suchenden unterschiedlich interpretiert. Ich kombiniere daher bekannte Verfahren zur Erkennung von Entwurfsmängeln auf der Basis von Metriken mit maschinellen Lernverfahren. Damit kann automatische Entwurfsmangelerkennung individuellen Sichtweisen angepasst werden.

Im folgenden Abschnitt 2 wird kurz auf Entwurfsmängel und deren Erkennung eingegangen. Eine ausführliche Beschreibung der Konzepte ist in [2] zu finden. Abschnitt 3 fasst Kriterien der Evaluation zusammen. Die abschließende Zusammenfassung verweist auf die Ergebnisse der Evaluation.

2 Adaptive Erkennung

Entwurfsmängel sind Programmeigenschaften, die auf potentiell fehlerhaften Entwurf eines Software-Systems hindeuten.

In der Literatur werden diese als *„Design Heuristics“* [4], *„Design Characteristics“* [7] oder *„Bad Smells“* [1] beschrieben. Die Autoren bezeichnen Entwurfsmängel i. d. R. durch Metaphern und erklären dem Software-Entwickler und Software-Architekten wie solche Entwurfsmängel erkannt und behoben werden können.

Um Entwurfsmängel automatisch zu erkennen, kombiniere ich objektorientierte Metriken (OO-Metriken) mit Verfahren des maschinellen Lernens.

Abbildung 1 zeigt das Grundkonzept zur automatischen und adaptiven Erkennung von Entwurfsmängeln.

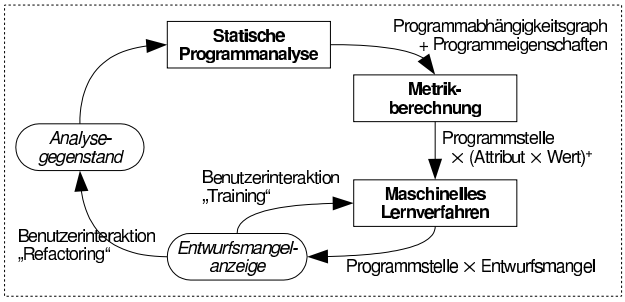

Abbildung 1: Grundkonzept zur adaptiven Erkennung von Entwurfsmängeln.

Ich folge dem Ansatz aus [3] und ordne jedem Entwurfsmangel eine Menge von Programmeigenschaften zu, die durch objektorientierte Entwurfsmetriken ausgedrückt werden. Häufig werden Größen-, Komplexitäts-, Kopplungs- oder Kohäsionsmetriken verwendet. Dieses Modell dient als Definition einer Strategie zur Erkennung jedes Entwurfsmangels.

Metriken werden durch statische Programmanalyse mit klassischen Methoden der Kontroll- und Datenflussanalyse sowie der abstrakten Interpretation ermittelt. Die Analyseergebnisse führen u. a. zu einem Programmabhängigkeitsgraphen, der als abstraktes Modell des Analysegegenstandes dient und zur Metrikberechnung genutzt wird.

Die Messwerte von Programmstellen werden an ein maschinelles Lernverfahren weitergeleitet. Dieses wird mit Beispielen von vorhandenen Entwurfsmängeln trainiert und konstruiert einen Entscheidungsbaum. Dieser entscheidet anhand der Messwerte, ob ein Entwurfsmangel vorliegt.

Abbildung 2 zeigt den aus zwei Phasen bestehenden Ansatz. In der Trainingsphase werden Programmstellen ausgewählt, für die einzeln manuell entschieden wird, ob der betrachtete Entwurfsmangel vorliegt. Die Messwerte dieser Programmstellen bilden — zusammen mit der Entwurfsmangelentscheidung — die Trainingsmenge, aus der sich ein initialer Entscheidungsbaum konstruieren läßt.

In der Erkennungsphase gibt der Anwender vor, welche Systemteile analysiert werden sollen. Hier wer-

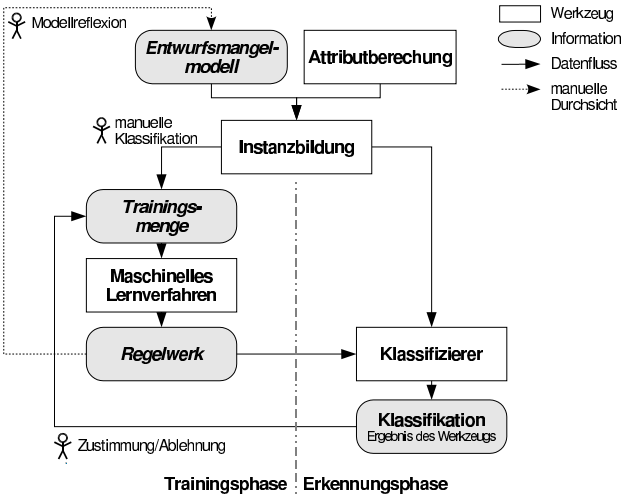

Abbildung 2: Lern- und Erkennungsphase mit Adaption des Entscheidungsbaumes und des Entwurfsmangelmodells.

den alle bisher unbekannten Programmstellen vermessen und alle Entscheidungsbäume der betrachteten Entwurfsmängel darauf angewandt. Jede Programmstelle, die als mangelbehaftet klassifiziert wurde, wird dem Anwender genannt.

Der Anwender prüft nun einzelne Fälle und akzeptiert den gefundenen Entwurfsmangel oder er lehnt ihn ab. In beiden Fällen kann die Programmstelle als weiteres Beispiel der Trainingsmenge hinzugefügt werden. Ignoriert der Anwender den Fund, bleibt die Trainingsmenge unverändert.

Trainings- und Erkennungsphase sind voneinander unabhängig. Auch während der Erkennung kann der Anwender weiterhin auffällige Programmstellen in die Trainingsmenge aufnehmen.

3 Kriterien der Evaluation

Eine Bewertung des Ansatzes ist nur durchführbar, wenn zu einer Menge von Programmstellen bereits bekannt ist, ob und welche Entwurfsmängel diese darstellen. Diese Information ist nur durch das Urteil von Testpersonen oder einer Expertengruppe zu erheben.

Durch eine Voruntersuchung werden daher Programmstellen von einer Expertengruppe manuell nach Entwurfsmängeln durchsucht. Die *Thurstone*-Methode [6] erlaubt die Einzelurteile zu akkumulieren. Man erhält eine „anerkannte“ Bewertungsvorgabe, die ein erstes Training des Verfahrens erlaubt.

Neben der Effizienz des Verfahrens, ist vor allem die Effektivität zu bewerten. Diese beschreibt die Erkennungsleistung des Verfahrens hinsichtlich der folgenden verkürzt dargestellten Kriterien:

Reproduktionsleistung Eine Menge von Programmstellen wird zusammen mit der Klassifizierung durch eine Testperson als Trainingsmenge verwendet. Anschließend läßt man das Verfahren in der gleichen Menge von Programmstellen suchen. Gemessen wird die Leistung des Lernverfahrens, das Gelernte wiederzugeben.

Transferleistung Soll beurteilt werden, wie gut das Lernverfahren das Konzept eines Entwurfsmangels erlernt hat, so verwendet man nur einen Teil der Trainingsmenge zum Lernen und beobachtet, ob ungelernte Programmstellen ebenfalls korrekt klassifiziert werden.

Trainingsleistung Das Entwurfsmangelmodell ist erst dann anzupassen, wenn ausgeschlossen werden kann, dass die Transferleistung nicht durch Erweiterung des Trainingsspektrums verbessert werden kann. Hierzu wird eine Effektanalyse eingesetzt, um die Vollständigkeit der Trainingsmenge zu bestimmen.

Die Leistungsarten können mit Methoden des *Information Retrieval* [5] ermittelt werden. Man findet hierzu standardisierte Verfahren, wie *Precision/Recall*- und *ROC*-Analyse, die den unabhängigen Vergleich mehrerer Systeme erlauben.

4 Zusammenfassung

Entwurfsmängel lassen sich durch Metrik-basierte Interpretationsverfahren erkennen. Unterschiedliche Sichtweisen auf Entwurfsmängel erfordern, dass ein System zur automatischen Erkennung, sich diesen anpasst. Hierzu werden maschinelle Lernverfahren eingesetzt.

In meinem Vortrag zeige ich Kriterien zur Evaluation des Verfahrens, den aktuellen Stand der empirischen Untersuchung und aktuelle Ergebnisse.

Literatur

[1] Martin Fowler, Kent Beck, John Brant, William Opdyke, and Don Roberts. *Refactoring: Improving the Design of Existing Code*. Addison Wesley, 1999.

[2] Jochen Kreimer. Adaptive Erkennung von Software-Entwurfsmängeln. *Informatik — Forschung und Entwicklung, Springer-Verlag*, erscheint 2005.

[3] Radu Marinescu. *Measurement and Quality in Object-Oriented Design*. PhD thesis, University of Timisoara, 2002.

[4] Arthur J. Riel. *Object-Oriented Design Heuristics*. Addison Wesley, 1996.

[5] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Tokio, 1983.

[6] L. L. Thurstone. A law of comparative judgement. *Psychological Review*, 34:273–286, 1927.

[7] Scott A. Whitmire. *Object Oriented Design Measurement*. John Wiley & Sons, Inc., 1997.