

Zusammenfassung

Zunehmend komplexe Softwaresysteme und agile Entwicklungsmethoden machen es immer schwieriger, die Code-Qualität eines Softwareprojektes zu kontrollieren. Deshalb wäre ein System zur Fehlervorhersage wünschenswert. Wir glauben, dass Machine Learning das Potenzial bietet, um ein solches System zu ermöglichen. Diese Bachelorarbeit hat zum Ziel, bestehende Ansätze mit Konzepten der Textanalyse, insbesondere N-Grams, zu erweitern. Ausserdem soll eine Grundlage für zukünftige Arbeiten zu diesem Themengebiet geschaffen werden. Dafür wurde ein umfassendes und modulares Toolset entwickelt. Dieses ist in der Lage, Git-Repositories beliebiger Java-Projekte zu analysieren. Die dabei extrahierten Daten können dann als Lernbasis für einen Machine-Learning-Algorithmus verwendet werden. Damit versuchen wir vorherzusagen, wie viele Bugfixes eine Dateiversion in den kommenden Monaten erfahren wird. Die implementierte Lösung zeigte, dass statistisch signifikante Zusammenhänge zwischen den genutzten Features und der Fehleranfälligkeit von Java-Dateien bestehen. Jedoch konnten die Resultate der Experimente den Nutzen der N-Grams nicht bestätigen.

Schlüsselwörter: Fehlervorhersage, Machine Learning, Regression, N-Grams, Repository Mining, Software Metrics, Feature Design