

Masterarbeit in Informatik/Wirtschaftsinformatik

Ausgangssituation

Tests einer Software sind ebenso essentieller Bestandteil eines Softwareprojekts wie die eigentliche Entwicklungsarbeit. Abgesehen von elementaren Unit Tests benötigen umfassende Tests mit höherer Aussagekraft Eingangsdaten, weil sie die Schnittstellen eines Programms zur Außenwelt mit einbeziehen. Solche Fixtures sind zwar ebenfalls Teil eines Softwareprojektrepositorys, dienen aber letztlich auch nur dazu, bestimmte, bekannte Fälle zu prüfen oder Regressionen zu vermeiden. Insbesondere stark datengetriebene Softwareprojekte, wie sie in der Implementierung von kommunikationslastigen Softwares zum elektronischen Datenaustausch (electronic data interchange, EDI (Nöcker 2002)) existieren, benötigen mehr als nur eine Reihe von Fixtures; Daten treiben maßgeblich das Verhalten der Software. Sogenannte Fuzzer (Böck 2015) belegen Programmvariablen mit beliebigen, rein zufälligen Werten, um durch den „Datenmüll“ Programme zum Absturz zu bringen und Fehler aufzuzeigen. Werkzeuge und Methoden zur Generierung von Testdaten, die sowohl syntaktisch korrekt sind, als auch möglichst viele Codepfade ansprechen, existieren bisher nicht.

Aufgabenstellung

Generative Adversarial Networks (GANs) sind eine Technik aus dem Bereich des unüberwachten Lernens, die zur Generierung realitätsnaher Daten erfolgreich genutzt wird (Goodfellow, et al. 2014). Bekanntestes Beispiel ist wohl das Gemälde „Edmond de Belamy“, das beim Auktionshaus Christie’s für 432.500 US-Dollar versteigert wurde (monopol 2018). GANs bestehen aus zwei Netzwerken, dem *Generator* und dem *Discriminator*, die ein Nullsummenspiel gegeneinander spielen: Der Generator erzeugt Daten, die der Diskriminator von einem realen Datensatz unterscheiden muss. Ziel des Generators ist, einen Datensatz zu generieren, den der Diskriminator nicht mehr von der Realität unterscheiden kann. Generator und Diskriminator trainieren sich so gegenseitig.

Die Methodik der GANs soll in dieser Thesis genutzt werden, um Testdaten für EDI-Softwares zu generieren. Basierend auf einem kleinen Satz realer, existierender Testdaten sollen GANs neue, syntaktisch korrekte und realitätsnahe Testdaten erzeugen, die bisher nicht generiert wurden.

An wen richtet sich die Themenstellung?

Die Arbeit richtet sich an Studierende der Informatik oder Wirtschaftsinformatik, die Interesse am systematischen Testen von Software haben und sich mit modernen Methodiken des Deep Learning auseinandersetzen wollen. Ein voriger Besuch der Vorlesung „Computational Intelligence“ ist wünschenswert.

Betreuung

Dr.-Ing. Eric MSP Veith | eric.veith@offis.de

Prof. Dr. Sebastian Lehnhoff | lehnhoff@offis.de

Literaturverzeichnis

Böck, Hanno. „Fuzzing with american fuzzy lop.“ *lwn.net*, 2015.

Goodfellow, Ian, et al. „Generative Adversarial Networks.“ *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014)*. 2014. 2672–2680.

monopol. „Christie's versteigert Werk eines Algorithmus.“ *monopol Magazin*, 2018.

Nöcker, Gregor. *Die beleglose Spedition*. LIT Verlag, 2002.

