

33 Wie soziale Netzwerke bei der Behebung von Softwarefehlern helfen

Mathematiker und Informatiker nutzen Methoden der künstlichen Intelligenz, um automatisch große Mengen von Fehlerberichten in großen Softwaresystemen zu kategorisieren. Graphentheorie, wie sie auch in sozialen Medien zum Einsatz kommt, hilft ihnen dabei, diese Aufgabe besonders effizient zu erledigen.

Bugtracking in Open Source Software Projekten

Software kommt in unserer modernen Gesellschaft fast überall zum Einsatz, sei es in unseren Smartphones, im Auto, bei Onlinediensten, oder in Produktionsanlagen von Unternehmen. Ihre Programmierung ist dabei meist so komplex, dass diese ein Team von Softwareentwicklerinnen und –entwicklern erfordert. Häufig handelt es sich hierbei um bezahlte Mitarbeiterinnen und Mitarbeiter einer Firma, welche die Software herstellt, testet und verkauft. Immer häufiger wird Software aber auch auf Grundlage einer sogenannten Open Source Lizenz von einer Community erstellt, die aus vielen Freiwilligen besteht, die sich über Online-Plattformen austauschen und mittels Internet-Technologien global zusammenarbeiten. Populäre Beispiele für Software, die so entstanden ist, sind der Browser Mozilla Firefox, das mobile Betriebssystem Android oder auch die Software, welche die Grundlage für die Online-Enzyklopädie Wikipedia bildet.

Da es in diesen sogenannten Open Source Software (OSS) Communities häufig keine bezahlten Testnutzer gibt, ist das Feedback der Nutzer für die Qualität des Endprodukts von besonderer Bedeutung. Hierzu betreiben fast alle Projekte sogenannte „Bugtracking“-Plattformen. Dieser Name beruht auf der Tatsache, dass man einen Fehler in einer Software in der Informatik auch als „Bug“, (englisch für „Insekt“, oder Käfer) nennt - ein Begriff, der aus einer Zeit stammt, als Computerprozessoren noch so groß waren, dass ein darin umherkrabbelnder Käfer tatsächlich einen Fehler auslösen konnte. Hat man als Nutzer einer Open Source Software nun ein Problem, welches man für einen Fehler hält, kann man einen sogenannten „Bugreport“, d.h. einen Fehlerbericht, schreiben. Darin beschreibt man möglichst genau, welche Version der Software man verwendet, wie sich der Fehler ausdrückt, wann und unter welchen Umständen dieser auftritt und welches Ergebnis man eigentlich erwartet hätte. Nun hofft man darauf, dass die Entwickler sich das Problem anschauen und möglichst bald eine Lösung anbieten.

Leider stellt die schiere Anzahl solcher Fehlerberichte die Entwickler besonders erfolgreicher Open Source Projekte vor große Herausforderungen. So mussten die Entwickler von Mozilla Firefox über einen Zeitraum von zehn Jahren mehr als 112,000 Fehlerberichte von Nutzern abarbeiten. Wenn die Bearbeitung jeder dieser Berichte nur 10 Minuten dauert, würde eine einzelne Person hierzu mehr als 2.300 Arbeitstage benötigen. Interessanterweise zeigt eine Analyse aber auch, dass mehr als 80 % der Fehlerberichte sich nicht auf einen Fehler in der Software zurückführen lassen, sondern bspw. auf einem Missverständnis der Nutzer oder einer Fehlbedienung beruhen. Nur ein kleiner

Teil dieser großen Anzahl von Fehlerberichten hilft also tatsächlich bei der Verbesserung der Software. Wenn es uns nun also gelingt möglichst schnell genau diese hilfreichen Fehlerberichte zu identifizieren, hilft das dabei Softwarefehler möglichst schnell und effizient zu beheben. Können wir Methoden der Mathematik einsetzen, um dieses Problem zu lösen?

Soziale Netzwerke in Open Source Projekten

Heutzutage kennt bereits jedes Kind soziale Medien wie TikTok, Twitter, Instagram oder Facebook. Auch wenn sich diese Plattformen in vielen Details unterscheiden, liegt doch allen ein gemeinsames Konzept zugrunde, welches in der Mathematik bereits vor fast 300 Jahren entwickelt wurde: ein soziales Netzwerk von Nutzern, welches in Form eines sogenannten Graphen dargestellt werden kann. Einen solchen Graphen kann man sich in etwa vorstellen wie Malen nach Zahlen, d.h. er besteht aus Punkten, den sogenannten Knoten, die über Linien, den sogenannten Kanten, paarweise miteinander verbunden sein können. Graphen aus Knoten und Kanten werden in vielen Bereichen der Wissenschaft eingesetzt, um komplexe Beziehungen zwischen verschiedenen Dingen darzustellen, bspw. Bücher in einer Bibliothek, die über Querverweise miteinander verbunden sind, Tierarten in einem Ökosystem, deren Überleben voneinander abhängt, oder Menschen, die in einem sozialen Netzwerk miteinander kommunizieren, bekannt oder befreundet sind. Die mathematische Beschreibung und Analyse solcher komplexen Graphen ist die Grundlage vieler aus dem Alltag bekannter Technologien. Beispielsweise liegt der Suche mit Suchmaschinen wie Google oder Bing eine Analyse von Verbindungen, sogenannter Hyperlinks, zwischen Milliarden von Webseiten zugrunde. Freundschaftsempfehlungen in sozialen Medien wie Twitter oder Instagram werden auf Grundlage der Verbindungen zu unseren Freunden, sowie deren Verbindungen untereinander berechnet. Und wenn wir Online einkaufen, beruhen die von Einkaufsplattformen vorgeschlagenen Produkte häufig auf einem Graphen, in dem diejenigen Produkte miteinander verbunden werden, die häufig zusammen bestellt werden.

Doch wie kann die mathematische Analyse solcher komplexen Graphen nun dabei helfen, die Flut von Fehlerberichten in Open Source Software Projekten zu beherrschen? Hierzu haben sich Forscherinnen und Forscher einen Trick ausgedacht, bei dem die Mitglieder einer Open Source Community in Form eines sozialen Netzwerks dargestellt werden. Wann immer zwei Personen über die jeweilige Kollaborationsplattform miteinander kommunizieren werden diese über eine Kante miteinander verbunden. Über die Zeit entsteht so ein großer komplexer Graph, der die Community von Freiwilligen eines Open Source Projekts darstellt und der es uns erlaubt diese mit mathematischen Methoden zu analysieren.

Wann immer ein Nutzer nun einen Fehlerbericht einsendet steht nun nicht nur dieser Fehlerbericht zur Verfügung, sondern wir kennen auch den Knoten im sozialen Netzwerk, der zum jeweiligen Nutzer gehört. Wie stark dieser Knoten mit anderen Nutzern vernetzt ist, zu welchen anderen Knoten diese Verbindungen bestehen, oder an welcher Position im Netzwerk sich diese befinden kann nun – neben den Daten des Feh-

lerberichtes – von einem maschinellen Lernverfahren genutzt werden, um automatisch diejenigen Fehlerberichte zu identifizieren, die tatsächlich auf einen wichtigen Fehler in der Software hindeuten. Die Idee dahinter ist, dass die Position einer Person innerhalb einer Open Source Community Rückschlüsse darüber zulässt, wie gut sich die jeweilige Person mit den Details der Software auskennt und wieviel Erfahrung diese darin hat, ein Problem so zu beschreiben, dass die Entwicklerinnen und Entwickler den zugrundeliegenden Fehler im Code finden können. Eine künstliche Intelligenz kann nun also lernen, aus welchen „Regionen“ des sozialen Netzwerks der Open Source Community besonders hilfreiche Fehlerberichte kommen.

Ein Test dieses Verfahrens in einem Datensatz mit Millionen von Fehlerberichten zu verschiedenen Open Source Projekten zeigt, dass diese Methode in mehr als 90 % aller Fälle richtige Vorhersagen macht, d.h. von 100 Fehlerberichten, die die Methode als potenziell hilfreich identifiziert hat resultieren mehr als 90 tatsächlich aus einem Fehler in der Software. Das Verfahren kann daher genutzt werden, um schnell diejenigen Fehlerberichte zu identifizieren, welche mit großer Wahrscheinlichkeit tatsächlich auf einen Fehler in der Software zurückzuführen sind. Diese können von den Entwicklern dann zuerst angeschaut werden, um Probleme möglichst schnell zu beheben. Dieses Verfahren, welches auf mathematischen Konzepten beruht, die mehrere Jahrhunderte alt sind, trägt also in unserer digitalen Gesellschaft dazu bei, Softwarefehler möglichst schnell zu finden und zu beheben.