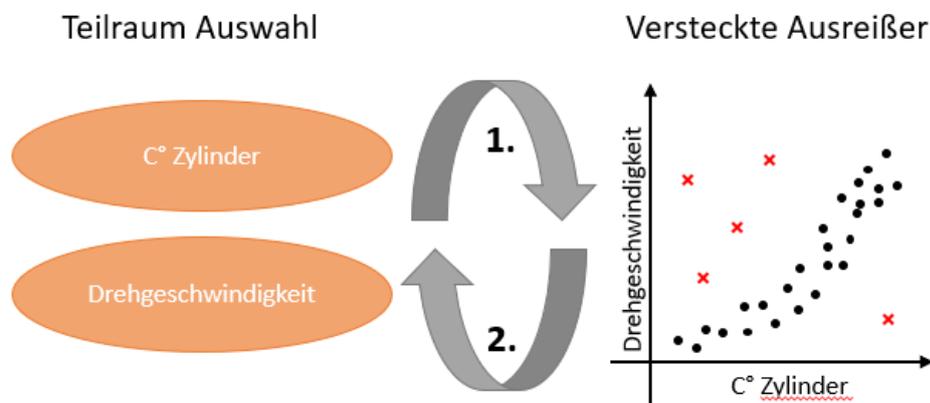


Robuste Teilraumsuche zur Vermeidung verheerender Fehler

Ausreißer sind Datenobjekte, die sich grundsätzlich von anderen (normalen) Datenobjekten unterscheiden. Dieser Unterschied ist in unterschiedlichen Teilräumen üblicherweise anders ausgeprägt. Betrachten wir zum Beispiel die Temperatur eines Zylinders und die Drehgeschwindigkeit eines Motors. Die Werte in diesen zwei Dimensionen hängen stark voneinander ab. Eine Beobachtung, dass die Temperaturen des Zylinders hoch, die Drehgeschwindigkeit aber gering, ist deshalb ein Ausreißer (s. b. die Abbildung). Wenn wir aber die Temperatur oder die Drehgeschwindigkeit jeweils für sich betrachten, ist die Beobachtung normal. Durch die Betrachtung dieser eindimensionalen Teilräume ist der Ausreißer also nicht zu erkennen. Das ist es, was man unter einem versteckten Ausreißer versteht: Der Ausreißer ist nur zu erkennen, wenn man Drehgeschwindigkeit und Temperatur zusammen betrachtet.

Analytische Untersuchungen in diesem Kontext sind notorisch schwierig. Das schon alleine aufgrund der großen Anzahl an Verfahren zum Erkennen von Ausreißern. Deshalb haben wir ein Verfahren entwickelt, das für einen Datensatz synthetisch Ausreißer generiert. Mithilfe dieses Verfahrens möchten wir nun Teilräume finden, die robust gegenüber versteckten Ausreißern sind, d. h. Teilräume, in denen Ausreißer nur schwer zu verstecken sind.

Um solche Teilräume zu erhalten, schlagen wir ein iteratives Vorgehen vor (vgl. Abbildung). Die zwei Schritte kommen wiederholt zur Anwendung:



1. Platzieren von synthetischen versteckten Ausreißern, basierend auf einer Auswahl von Teilräumen
2. Anpassung der Auswahl an Teilräumen, basierend auf den versteckten Ausreißern

Eine positive Konvergenz des Prozesses bedeutet, dass es schwierig bis unmöglich ist, versteckte Ausreißer zu platzieren. Die dabei gefundenen Teilräume sind also robust gegenüber versteckten Ausreißern. Natürlich hängen eine solche Konvergenz und auch die Ergebnisqualität stark von der genauen Vorgehensweise in 1 und 2 ab. Ziel dieser Arbeit ist es, Alternativen für 1 und 2 zu entwickeln und zu vergleichen. Bei der Bearbeitung dieses Themas entwickeln Sie vielfältige Fähigkeiten im Bereich der Datenanalyse, sowohl hinsichtlich Algorithmenentwurf als auch Evaluation.

Ansprechpartner

Georg Steinbuss, M. Sc.
Am Fasanengarten 5

georg.steinbuss@kit.edu
76131 Karlsruhe

+49 721 608-42757
Gebäude: 50.34

Raum: 362