DER STANDARD

Der Standard
21-08-2019
Seite: **13**
Land: **Österreich**
Region: **Überregional**

Auflage: **71.654**
Reichweite: **525000**
Artikelfläche: **38813 mm²**
Skalierung: **67%**

Wenn die Lücken im Datenmaterial ein Muster ergeben

Ein Wiener Forschungsprojekt arbeitet an der visuellen Aufbereitung der Datenlandschaften hinter medizinischen Langzeituntersuchungen

Medizinische Langzeituntersuchungen beschäftigen sich oft mit Zusammenhängen von Risikofaktoren und dem Auftreten von Krankheiten. In diesen epidemiologischen Studien versuchen Wissenschafter beispielsweise herauszufinden, welche Umwelteinflüsse, Verhaltensmuster oder Vorprägungen zu Krebserkrankungen führen. Oft werden Daten tausender Studienteilnehmer, die über mehrere Jahre hinweg beobachtet werden, zusammengeführt. Zur Vielfalt dieser Datensets gehören auch ihre Lücken: Leerstellen in den Datenbanken, die aus sehr unterschiedlichen Gründen entstehen können.

Um Erkenntnisse nicht zu verfälschen, muss man sich dieser Leerstellen bewusst sein. Datensätze müssen ergänzt, gestrichen oder korrigiert werden. Softwaresysteme können dabei wertvolle Beiträge leisten. Am Wiener Zentrum für Virtual Reality und Visualisierung (VRVis) versucht man den Wissenschaftern mit einem Visual-Analytics-Ansatz unter die Arme zu greifen. Mit dem hier entwickelten Werkzeug Vivid („Visual Analysis of Missing Values in Cohort Study Data") können Daten – und ihre Lücken – visuell aufbereitet, vervollständigt und überprüft werden.

Shiva Alemzadeh entwickelt das System am VRVis in Kooperation mit der Universität Greifswald und der Universität Magdeburg, wo sie ihr Doktorrat im Bereich Data-Science macht. In ihrer Forschungsarbeit beschäftigt sie sich mit den verschiedenen Charakteristiken fehlender Daten. Da gibt es zum einen Lücken, die in einer Abhängigkeit zu weiteren Faktoren innerhalb der Datensätze stehen. „Das ist etwa der Fall, wenn Blutdruckwerte nur bei älteren Patienten ausgewiesen sind – wo also eher Probleme zu erwarten sind –, nicht aber bei jüngeren", gibt die Forscherin ein Beispiel.

Zum anderen gibt es auch Leerstellen, bei denen zwar eine Abhängigkeit vorhanden ist, diese aber außerhalb der Daten liegt. Alemzadeh: „Wenn eine Patientin nicht zur Untersuchung für eine Migränestudie erscheint, könnte es etwa sein, dass eben eine Migräne selbst der Anlass dafür ist." Und zuletzt gibt es jene Lücken, die keinerlei Systematik unterliegen. Der Patient hat den Termin vielleicht nur übersehen oder der Arzt vergessen, einen Wert einzutragen.

Gesundheitsmaßnahmen sollten auf gutem Datenmaterial fußen.
Foto: Getty / iStock / George Rudy

Die visuelle Aufbereitung der Daten, ihrer verschiedenen Kategorien und Abhängigkeiten soll Muster in den Leerstellen vor Augen führen. Gleichzeitig übernimmt Alemzadehs Werkzeug die Ergänzung von Daten. Leerstellen werden durch statistische Verfahren – man spricht von Imputation – mit wahrscheinlichen Werten aufgefüllt, um Verzerrungen zu verringern. Zuletzt werden im Zuge einer Gültigkeitsüberprüfung die vorhergesagten Werte noch auf ihre Plausibilität geprüft.

Das Vivid-Framework wurde auf die Kategorien und Variablen einer epidemiologischen Studie zugeschnitten. In Zukunft soll der Nutzerkreis nicht nur in der medizinischen Forschung ausgeweitet werden. Alemzadehs glaubt, dass das Werkzeug durchaus auch in anderen Bereichen der empirischen Forschung Anwendung finden kann. „Wir hoffen, dass wir den Ansatz generalisieren und beispielsweise auch in politischen Umfragen anwenden können, um die Qualität der Daten zu garantieren", sagt die Entwicklerin. „Lücken in den Datensätzen sind ein Problem, das in allen Forschungsbereichen auftaucht." (pum)