

Psychologie

Big Data und Machine Learning in der Psychologie

Vom US-Präsidentenwahlkampf zu seriöser wissenschaftlicher Forschung: Die Speicherung und Auswertung großer Datenmengen (Big Data) sind in der Psychologie ein topaktuelles Thema. Wie generiert man die Daten, und was ist im Hintergrund an IT-Ausstattung und -Kenntnissen nötig, damit die Forscher effektiv und verantwortungsvoll arbeiten können?

VON CAROLINE ZYGAR UND JOHANNES ALBERT-VON DER GÖNNA

SPÄTESTENS SEITDEM bekannt wurde, dass für den amerikanischen Präsidentenwahlkampf offenbar Daten sozialer Netzwerke ausgewertet wurden, um Persönlichkeitsprofile der Wählerinnen und Wähler zu erstellen, ist klar: Die Speicherung großer Datenmengen und deren Verarbeitung ist auch in der Psychologie ein aktuelles und besonders relevantes Thema. Entsprechende Fragestellungen und die Weiterentwicklung von Forschungsmethoden eröffnen neue Perspektiven auf das menschliche Verhalten und Erleben. Gleichzeitig wird aber auch deutlich, dass ein verantwortungsvoller Umgang mit diesen Daten dringend notwendig ist.

In der psychologischen Forschung sind bisher vornehmlich Studienansätze üblich, die einmalig Daten von Personen erheben: Entweder werden Studienteilnehmerinnen und -teilnehmer eingeladen, um ihr Verhalten und ihre Reaktionen in einer bestimmten Situation zu beobachten, oder ihnen werden Fragebögen vorgelegt, in denen sie Aufgaben zu lösen (z. B. Intelligenztest) oder eine eigene Einschätzung zu

treffen haben (z. B. Persönlichkeitsfragebogen). Dieses Vorgehen führt typischerweise nicht nur zu recht kleinen Datensätzen, da lediglich eine begrenzte Anzahl an Personen beobachtet werden kann, auch stellen diese Daten in der Regel nur Momentaufnahmen dar. Mögliche Veränderungen werden nicht erfasst, und die Übertragbarkeit dieser Beobachtungen auf andere Situationen bleibt oft ungeklärt.

Forschungsdaten sammeln mit dem Smartphone

Mit der rasanten Verbreitung von Smartphones hat sich die Studienlandschaft verändert: Als vielseitiger Alltagsbegleiter bieten sie auch die Möglichkeit, umfassende Forschungsdatensätze zu erzeugen. Immer mehr Personen können über ihr eigenes Smartphone kontinuierlich befragt



Studierende erproben die Datenerhebung am Smartphone im empirisch-psychologischen Praktikum.

werden – ein Vorgehen, das unter dem Begriff „Experience Sampling“ bekannt ist. Auf jedem Gerät können zudem vielfältige Nutzungsdaten aufgezeichnet werden, so etwa die Anzahl der Anrufe, die verwendeten Apps oder Sensordaten wie etwa die GPS-Position. Dies bezeichnet man als „Data Logging“. Das Einverständnis der betroffenen Personen vorausgesetzt, können diese mit Hilfe des Smartphones generierten Datenmengen ausgewertet werden, um psychologische Fragestellungen zu beantworten.

DIE AUTOREN

Caroline Zygara, M. Sc. ist wissenschaftliche Mitarbeiterin am Lehrstuhl für Psychologische Methodenlehre und Diagnostik der LMU München.

Dr. Johannes Albert-von der Gönna ist wissenschaftlicher Mitarbeiter am Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften.

Untersuchung von Partnerschaften ...

Am Lehrstuhl für Psychologische Methodenlehre und Diagnostik der LMU München wird das Smartphone auf diese Weise eingesetzt. Ein Forschungsprojekt unter der Leitung von Felix Schönbrodt etwa nutzt Experience Sampling zur Untersuchung von Partnerschaften. Teilnehmende Paare erhalten eine App, über die sie tagsüber in regelmäßigen Abständen zu den Bedürfnissen in ihrer Partnerschaft befragt werden: Was wünschen sie sich von ihren Partnern? Wie verändern sich diese Bedürfnisse? Welche Rolle spielt das eigene Verhalten und das des Partners? Die wiederholte Befragung über mehrere Wochen hinweg ergibt einen Datensatz, der viel größer als herkömmliche Datensätze in der Psychologie ist und der es erlaubt, Veränderungen im Zeitverlauf zu beobachten und abzubilden.

... oder von Persönlichkeitsunterschieden

Auch zum Thema Persönlichkeit und Verhalten wird an der Lehrereinheit geforscht. Hierzu wird mittels Data Logging konkretes Verhalten am Smartphone aufgezeichnet. Die Forschergruppe um Clemens Stachl geht der Frage nach, inwieweit sich Persönlichkeitsunterschiede in der Nutzung des Smartphones spiegeln. Bereits im Rahmen seiner

Doktorarbeit untersuchte er die (anonymisierten) Nutzungsdaten von über 100 Personen und fand erste Hinweise darauf, dass Persönlichkeitseigenschaften die Nutzung unterschiedlicher App-Kategorien auf dem Smartphone vorhersagen können.

Was ist aus Sicht der IT nötig?

Als umfassender IT-Dienstleister stellt das Leibniz-Rechenzentrum (LRZ) die Infrastruktur für den Betrieb der Server

zur Verfügung, die für solche Datenerhebungen am Smartphone notwendig sind. Gleichzeitig erfordert die Vielzahl der aufgezeichneten Aktionen und (Sensor-)Informationen moderne, computergestützte Auswertungsverfahren, um aus den großen Datenmengen relevante Informationen herauszufiltern. Diese Methoden werden unter dem Begriff „Machine Learning“ zusammengefasst. Auch für deren Einsatz wird zunehmend Rechenleistung benötigt, die von normalen Arbeitsplatzcomputern in vertretbarer Zeit nicht mehr aufgebracht werden kann. Das LRZ bietet den Wissenschaftlerinnen und Wissenschaftlern optimale Voraussetzungen und fachkundige Unterstützung, um solche Analysen auf parallelen Cluster- und spezialisierten Machine Learning-Systemen umzusetzen.

Neben der Notwendigkeit, diese neuen Analysemethoden zu beherrschen und anzuwenden, entstehen auch neue Herausforderungen im Datenmanagement: Um Forschungsergebnisse für andere Personen reproduzierbar und erweiterbar zu machen, ist es erstrebenswert, die Daten öffentlich zur Verfügung zu stellen. Hierbei muss an erster Stelle der Schutz personenbezogener Daten stehen bzw. die Anonymität der Studienteilnehmerinnen und -teilnehmer gewahrt werden. Zudem ist eine nachvollziehbare Dokumentation notwendig, um die Daten später sinnvoll auswerten zu können. Hierzu müssen den Forschenden weitere neue Kompetenzen vermittelt werden, da eine nachhaltige Aufbereitung und Dokumentation bei ansteigender Datenmenge zunehmend anspruchsvoller und zeitaufwändiger wird. Das Streben nach Open Science unter Wahrung der Persönlichkeitsrechte von Studienteilnehmerinnen und -teilnehmern wird an der LMU München durch das interdisziplinäre Open Science Center unterstützt.

Auswirkungen von Big Data auf die Lehre

Damit der Nachwuchs an Psychologinnen und Psychologen von den neuen Entwicklungen nicht abgehängt wird, muss die universitäre Lehre angepasst werden. Lehrstuhlinhaber Markus Bühner macht deutlich: „Big Data ist ein wichtiges Thema für die Psychologie und muss daher nach und nach in die Lehre integriert werden. Zukünftig sollten im Rahmencurriculum Seminare zum Thema Machine Learning angeboten werden, damit das Fach mit den neuen Möglichkeiten wächst.“ An seinem Lehrstuhl wurde in diesem Wintersemester zum ersten Mal ein empirisch-psychologisches Praktikum für Studierende im Bachelorstudiengang Psychologie angeboten, das sich mit dem Thema Data Logging in Kombination mit Open Science beschäftigte. Zeitgleich behandelte Lehrstuhlmitarbeiter Florian Pargent in einer Vorlesung für Masterstudierende der Psychologie erstmals das Thema Prädiktive Modellierung mit Techniken des Machine Learning.

Die Psychologie ist selbstverständlich nur einer unter vielen Fachbereichen, die sich an diesen Entwicklungen beteiligen. Insbesondere in der Statistik und Informatik werden die Methoden erarbeitet, die derartig komplexe Analysen weiter optimieren und damit die Chancen von Big Data noch besser nutzbar machen. In Zukunft wird es bei den beschriebenen Formen psychologischer Forschung deshalb noch mehr auf die Zusammenarbeit mit technischen Disziplinen und Dienstleistern wie dem Leibniz-Rechenzentrum ankommen, um die Erfahrung und das Können der verschiedenen Fachrichtungen zu verknüpfen und damit die wissenschaftliche Forschung und daraus resultierende Anwendungen gemeinsam und verantwortungsvoll weiterzuentwickeln. ■



WWW

www.esm.researchcloud.de – Forschungsprojekt an der LMU München: Nähe und Distanz in Partnerschaften

www.osc.uni-muenchen.de – Open Science Center der LMU München