

Software-Entwicklungsprojekt zum Data Mining in der Medizin

Die Suche nach verborgenen Schätzen

In der Medizin fallen ungeheure Datenmengen an, die voller potenziell interessanter Informationen stecken. Menschen sind oft nicht mehr in der Lage, diese „Schätze“ zu heben, aber der Computer kann die nötige Vorarbeit leisten, um den Blick des Experten auf das Wesentliche zu lenken.

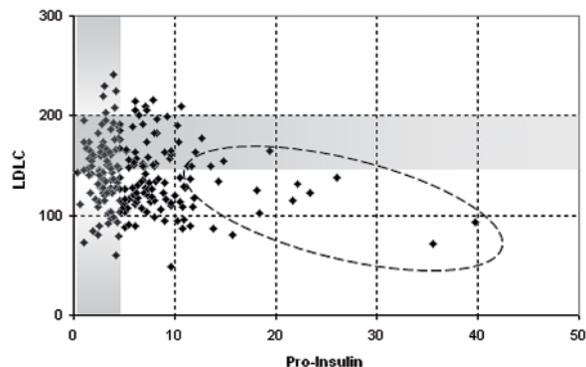
Es war ein klassischer Fall, wie er an Universitätskliniken häufig vorkommt: Ein Sportmediziner aus Freiburg und ein Laborarzt aus München untersuchten, wie körperliche Aktivität bestimmte Laborwerte aus dem Formenkreis des Metabolischen Syndroms (s. a. S. 224) beeinflusst. Das erhoffte Ergebnis blieb jedoch aus und deshalb verschwanden die an über 400 Probanden erhobenen Daten in der Schublade. Fast 5.000 Werte waren damit trotz des hohen Studienaufwands verloren, weil sich niemand fand, um die große Datenmenge in vertretbarer Zeit auf andere womöglich interessante Ergebnisse zu durchforsten.

Zufällig kam im Rahmen eines völlig unabhängigen Forschungsprojekts der Trillium GmbH die Sprache auf diese erfolglose Studie. Da gerade ein Testdatensatz für den Prototypen eines neuen Computer-Programms gesucht wurde, unterzogen wir die anonymisierten Resultate mit so genanntem Data Mining einer Nachprüfung.

Unter diesem sehr aktuellen Begriff fasst man eine Vielzahl von Techniken zusammen, die mit Computertechnik nach verborgenen

Strukturen und Zusammenhängen in großen Datensätzen suchen. Dazu gehören zum Beispiel bioinformatische Auswerteverfahren für die nebenstehend beschriebenen Gensequenzierungen. Auch bei so praktischen Anwendungen wie der Videoüberwachung auf Bahnhöfen oder der Analyse des Wählerverhaltens leisten sie wertvolle Dienste. Die Programme sind keineswegs intelligenter als Menschen, können jedoch in tausend- oder millionenfachen Rechenzyklen schlicht „alles mit allem“ vergleichen, um am Ende den Blick des Experten auf potenziell interessante Ergebnisse zu lenken.

Im vorliegenden Fall des missglückten Experiments förderte das Programm eine ganze Reihe bekannter Zusammenhänge wie Blutzucker und HbA1c zutage. In



Aus 5.000 Werten (Alter, BMI, Blutdruck, Labordaten) filterte der Computer eine Subgruppe mit extrem hohen Proinsulin- und sehr niedrigen LDL-Cholesterin-Spiegeln heraus. Die Referenzbereiche sind grau hinterlegt (Quelle Hoffmann G et al. *Klin Chem Mitteilungen* 2010; 41:160-164. Download: www.dgkl.de, Förderprojekt der Stiftung der DGKL e. V.).

einem kleinen Subkollektiv entdeckte es aber auch die in der Abbildung dargestellte, bislang nicht beschriebene Beziehung zwischen LDL-Cholesterin und Proinsulin. Ob dieser Befund reiner Zufall ist oder klinische Bedeutung hat, können nur gezielte weitere Studien klären.

Die Algorithmen für das Data Mining sind seit langem bekannt und allgemein zugänglich. Entscheidend für den Erfolg eines solchen Projekts ist das Zusammentragen und Filtern der Rohdaten aus externen Quellen und eine fachgerechte Aufbereitung. Da hierfür rund 90 Prozent der gesamten Arbeit anfallen, werden im Rahmen unseres Forschungsprojekts Verfahren entwickelt, die diese Schritte weitgehend automatisieren sollen. ❁

gh



Die Dr. Neumann & Kindler Ltd. & Co. KG (labcore®) unterzeichnete kurz vor Weihnachten mit der Trillium GmbH eine Vereinbarung zur Kooperation im Bereich Data Mining. Ein Schwerpunkt wird die Entwicklung von Schnittstellen zur Datenübernahme aus externen Datenquellen sein.