

IT hilft helfen – Data Mining in der medizinischen Forschung

Wer Leben retten will, darf keine Zeit verlieren. Der Einsatz moderner Data-Mining-Systeme senkt nicht nur die Kosten der Entwicklung wirksamer Behandlungsmethoden, sondern beschleunigt die Forschung. **VON JOSEF SCHMID***



Jahr für Jahr erkranken allein in den USA fast 3000 Kinder an einem Gehirntumor. Beinahe die Hälfte von ihnen wird in-nerst fünf Jahren an der für Kinder bösartigsten Krebsform sterben. Die anderen werden mit den Spätfolgen leben müssen. So nüch-tern Zahlen sind, hinter jeder verbergen sich individuelle Schicksale mit einer individuel- len Krankengeschichte, einem individuellen Krankheitsverlauf und einer individuellen Behandlungsstrategie. Denn insbesondere bei pädiatrischen Hirntumorerkrankungen sind allgemein gesicherte Erkenntnisse für eine wirksame Behandlung rar. Die behan- delnden Ärzte müssen neue Therapieansätze erproben und die Ergebnisse dokumentie- ren, damit geeignete Behandlungsstrategien entwickelt werden können. Auf diese Weise

werden jährlich tausende von Patientendos- siers geführt und zigtausende Fachartikel zur Charakterisierung, Diagnose und Be- handlung bösartiger Tumore publiziert. Und in allen sind wertvolle Informationen versteckt – unzählige Puzzleteile, um den Kampf mit dieser heimtückischen Krankheit aufzunehmen.

Am Children's Memorial Research Cen- ter in Chicago, einem führenden Kinderspi- tal und Forschungsinstitut der USA, hat sich Dr. Eric Bremer, Direktor des Brain Tumor Research Program, eine gewaltige Aufgabe vorgenommen: Durch systematische Ana- lyse der Patientendaten und der medizini- schen Forschungsliteratur klassifiziert der Wissenschaftler Tumorarten, erkennt Risi- kofaktoren und identifiziert Erfolg verspre-

chende Behandlungsstrategien, die zuvor in der grossen Menge der Informationen übersehen wurden.

Den Feind kennen

Eines seiner vorrangigen Ziele ist es, eine Da- tenbank mit den diversen Gen-Expressionen von Gehirntumoren bei Kindern aufzubauen. Deren Daten sollen mit den Ergebnissen frü- herer und laufender Forschungen korreliert werden. Grundvoraussetzung für eine ge- zielte Behandlung ist die exakte Klassifizie- rung des jeweils auftretenden Tumors. Die Mehrzahl der Gehirntumore bei Kindern kann zwölf Kategorien mit Unterkategorien

* Josef Schmid ist Director Consulting bei SPSS Schweiz.

zugeordnet werden: Glioma, Ependyom oder Medulloblastoma heisst der Feind. Oft erfolgt die Klassifizierung subjektiv, da sich die Tumorformen nur schwer nach ihrem äusseren Erscheinungsbild unterscheiden lassen. Zudem sind nur wenige verlässliche Tumormarker bekannt, die eine eindeutige Identifikation zulassen.

Ausser der zuverlässigen Zuordnung der Krebsart ist es für den Behandlungserfolg entscheidend, das exakte Entwicklungsstadium festzustellen. Und so schwer eine eindeutige Diagnose ist, so bedeutsam ist sie für den geeigneten Therapieansatz. Eine falsche Zuordnung kann drastische Folgen



ILLUSTRATION: CW/THU

für die kleinen Patienten haben: Wird fälschlich ein fortgeschrittenes Stadium diagnostiziert, wird auch eine unnötig aggressive Behandlung gewählt – etwa eine Chemotherapie, die nicht nur erhebliche unmittelbare Beschwerden, sondern auch Dauerschädigungen verursachen kann.

Versteckt in Datenbergen

Der Schlüssel zu einer eindeutigen Zuordnung von Art und Stadium eines Tumors liegt in dessen Molekularstruktur. Mit so genannten Microarray-Experimenten untersucht Bremer die DNA-Sequenzen und ermittelt mit einer Data-Mining-Anwendung das individuelle Muster der tumortypischen Gen-Expression. Pro Probe kann dies 7000 bis 30 000 Variablen ergeben. Erst

mit leistungsfähigen Auswertungsmethoden erhalten diese riesigen Datenmengen einen Sinn.

Bremer hat bei der Microarray-Datenanalyse zahlreiche Probleme zu lösen, da einer geringen Zahl von Samples eine grosse Zahl von Genen gegenübersteht. So weist der Datensatz von Bremer gerade mal 133 Proben, aber fast 7000 Variablen auf. Für verwertbare Ergebnisse soll das Data-Mining-Programm direkt auf die Daten zugreifen und diese gemäss den Anforderungen aufbereiten. Dabei ist auch zu gewährleisten, dass der jeweiligen Genausprägung tatsächlich eine biologische Variation und nicht eine Verunreinigung während der Präparation des Samples zugrunde liegt. Dies lässt sich mit einer Analyse-Software sicherstellen, die mit den jeweiligen Proben auch die Qualitätsparameter verarbeitet.

In einer ersten Testphase verknüpfte Bremer seine eigenen mit öffentlich zugänglichen Forschungsdaten. Mit den exakt beschriebenen Vergleichswerten liess sich die Zuverlässigkeit des Systems für die Zuordnung schwer klassifizierbarer Gehirntumore überprüfen. Mit einer 95-prozentigen Reliabilität ist das System sogar für die Zuordnung auf Unterkategorien geeignet. Wurden zuvor fünf unterschiedliche Anwendungen für weniger präzise Microarray-Analysen benötigt, unterstützt nun eine Anwendung sämtliche Analyseschritte. Zwei Vorhersagemodelle liefern ergänzende Daten für die zuverlässige Klassifikation: Während neuronale Netze genauere Ergebnisse hervorbringen, zeigen Entscheidungsbäume, wie die Tumoren klassifiziert werden und enthüllen sogar potenzielle Tumormarker. Sobald diese Marker validiert sind, können auch Labore ohne Microarray-Technologie mit diesen Informationen Antikörper für neue Diagnosemethoden entwickeln.

Doch ist die zweifelsfreie Klassifizierung einer Krebsform nur ein erster Schritt auf dem Weg zu wirksamen Behandlungsmethoden. Für weitergehende Erkenntnisse müssen die Ergebnisse um klinische, pathologische und biochemische Informationen ergänzt werden. Schliesslich zeigt die Kombination mit Patientendaten, welche Therapien die grössten Erfolge bei der Bekämpfung pädiatrischer Gehirntumore zeitigen. An dieser Stelle gelingt der Sprung von der Bioinformatik, die vorwiegend Daten erfasst und organisiert, zur biomedizinischen Informatik: Informatiksysteme ermögli-

chen neue Verfahren der Diagnostik, Therapie und Medikamentenentwicklung. Insbesondere der automatisierten Informationserfassung aus der biomedizinischen Literatur und der textbasierten Informationsverarbeitung wird dabei eine bedeutende Rolle zukommen.

Auch Bremer nutzt ein System, um relevante genetische Informationen in Volltexten aufzuspüren und diese mit den Forschungsergebnissen der

pädiatrischen Krebsbehandlung zu korrelieren. Die Lösung stellte er in Zürich an der Life Sciences Grid Conference 2005 vor: Eine flexible Text-Mining-Applikation in einer Grid-Computing-Umgebung kombiniert mit einer Anwendung für Predictive Analytics. So sind prädiktive Anwendungen nicht nur für betriebswirtschaftliche Zwecke einsetzbar, um zum Beispiel Kaufverhalten oder die Wahrscheinlichkeit einer Vertragskündigung vorherzusagen, sondern auch für die Analyse von Gen-Expressionen: «Business Intelligence» wird «Health Intelligence».

Wissen für alle

Gerade im Austausch von Daten und Wissen sieht Bremer einen kritischen Erfolgsfaktor für die Erforschung und Behandlung von Gehirntumoren. Aus diesem Grund will er seine Analysetechnologie über einen Web-Server auch anderen Wissenschaftlern zugänglich machen: Jeder, der Gehirntumore von Kindern erforscht, soll die Ergebnisse seiner Microarray-Experimente einspeisen können. Während die Forscher verlässliche Analysen und Vorhersagen erhalten, werden im Gegenzug die zugrunde liegenden Datensätze ergänzt und aktualisiert. Denn eines ist klar: Je mehr und genauere Informationen die Datenbank enthält, umso wertvoller werden die Ergebnisse sein und umso mehr Leben werden gerettet. So sollen Data und Text Mining helfen, eine grosse Zahl an Genen zu entdecken, die die Entwicklung von Tumoren beeinflussen. Bei allem ist sich Bremer bewusst, dass es nicht den einen richtigen Weg gibt, um Daten zu analysieren und Erkenntnisse zu gewinnen. Auf dem leidvollen Weg zum besten Ergebnis wird er verschiedene Methoden ausprobieren müssen. Erfahrung, Ausdauer, Frustrationstoleranz und leistungsfähige, flexibel einsetzbare Technologien sind dabei unabdingbar. Doch Bremer hat ein Ziel: Er will die «Prime Drug Targets», also die Zielmoleküle für die optimale Wirkung neuer Behandlungsformen, und damit die Basis für die Heilung von Krebs finden. IT wird ihm dabei helfen. ■

Data-Mining-Werkzeuge leisten in der Krebsforschung wichtige Dienste.