

Bioinformatik und Bio Data Science – ein Vergleich

Wie man biologische Daten analysiert

Im Rahmen der Vorbereitungen für die Serie „Life Science Career Paths“ stieß die ÖGMBT auf ein Berufsfeld, das heiß begehrt ist. Was die Bio Data Science macht, wo man das lernen kann und was sie von der Bioinformatik unterscheidet.

Als Anfang des neuen Jahrtausends erstmals ein vollständiges menschliches Genom sequenziert wurde, hatte das sogenannte „Shotgun Sequencing“ einen entscheidenden Anteil daran. Dabei werden lange DNA-Ketten zufällig fragmentiert, die Fragmente sequenziert und anschließend versucht, die Puzzlestücke zu einem großen Ganzen zusammenzuführen. Die dafür erforderlichen Algorithmen stellte eine Fachrichtung namens Bioinformatik zur Verfügung, die damals zum ersten Mal ins Licht der Öffentlichkeit trat.

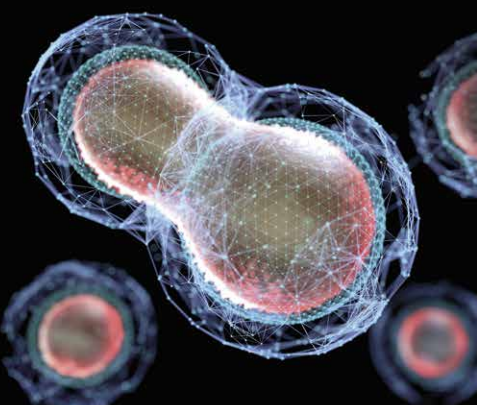
Heute ist die Bioinformatik längst eine selbstständige Teildisziplin der Biologie, und ihre Anwendungen beschränken sich keineswegs auf die Vervollständigung von Genomen. In der Erforschung von Proteinen, RNA oder Stoffwechselprodukten – oder allgemeiner: in der Systembiologie – treten ebenso große Datenmengen auf, die nach spezieller, eigens dafür geschaffener Software rufen. „Wissenschaftler, die in diesem Bereich arbeiten, sind größtenteils mit Methodenentwicklung beschäftigt, für die man ein tiefes Verständnis der biologischen Zusammenhänge benötigt“, sagt Thomas Rattei, Professor für Computational Biology an der Universität Wien und Vorstandsmitglied der ÖGMBT.

Neben der Bioinformatik ist in den vergangenen Jahren aber auch ein Aufgabenfeld entstanden, für das der Ausdruck Bio Data Science gebräuchlich wurde und das sich mit der Analyse von Daten biowissenschaftlichen Ursprungs beschäftigt. Im Grunde stammt diese Fachrichtung von der Statistik erhobener Messwerte ab, die in quantitativ arbeitenden Wissenschaften seit langem zur methodischen Basisausstattung gehört. Doch mit der Menge der Daten, die erzeugt, und neu verfügbaren Computer-unterstützten Methoden hat sich das Repertoire an Werkzeugen stark verbreitert und die neue Bezeichnung „Data Science“ nahegelegt. „Vieles davon ist klassische Statistik, etwa die Verwendung multivariater Ansätze. Dazu kommt der Einsatz von Modellen, z. B. aus dem Bereich Machine Learning“, sagt dazu Michael Wildauer, der ein Team von Data Scientists am Wiener Standort des Pharmaunternehmens Takeda leitet, dessen Aufgaben er vor kurzem

im Rahmen der „Career Path“-Serie der ÖGMBT vorstellte.

Statistik mit dem gewissen Etwas

Wo man im Speziellen von Bio Data Science spricht, kommen Methoden verschiedenen Ursprungs zusammen, wie Rattei erklärt: „Ansätze aus dem Bereich des maschinellen Lernens können auf die verschiedensten Fragestellungen angewendet werden. Dem stehen Methoden wie die Genomanalyse gegenüber, die in der Bioinformatik entwickelt wurden und nun



Ein Beispiel: Doktoratsprogramm PAIR

Manche Hoffnung der Krebstherapie ruht auf der Bestrahlung mit Ionenstrahlen. Um deren Wirkung auf Krebsgewebe besser zu verstehen, haben sich FH Wiener Neustadt, Meduni Wien und Vetmed zu dem vom FWF geförderten Doktoratsprogramm „PAIR – Pre-clinical Ion Beam Research“ zusammengetan. Gemeinsam untersucht man präklinische Tumormodelle und bringt dazu Expertise aus Strahlenbiologie, medizinischer Bildgebung und Computerwissenschaften zusammen. Der zuletzt genannte Aspekt ist dabei die Domäne von Milica Kronic, der Leiterin des Studiengangs Bio Data Science an der FH Wiener Neustadt: „Für unsere Studierenden ergeben sich hier ganz neue berufliche Perspektiven und Karrierewege.“

routinemäßig von Bio Data Scientists angewendet werden.“

Nicht nur die Biowissenschaft, auch die Biotechnologie ist eine gute Quelle für jenes Datenmaterial, das von Data Scientists bearbeitet wird. Die beim ÖGMBT-Firmenmitglied Takeda tätige Experten analysieren etwa Daten, die aus dem ERP-System des Unternehmens stammen, oder solche, die das Qualitätsmanagement zur Verfügung stellt. Ergänzt wird dies durch beobachtete Auffälligkeiten, etwa Gebrechen der Anlage oder Fehler eines Operators. Zwei Arten von Fragestellungen können anhand der Daten aus diesen Quellen bearbeitet werden: „Zum einen machen wir explorative Analysen, bei denen unsere Produktionsprozesse ohne ausschlaggebenden Grund auf Verbesserungspotenzial untersucht werden. Wenn andererseits Abweichungen von den vorgegebenen Kriterien auftreten, geht es darum, herauszufinden, was passiert ist und wie man korrigierend eingreifen könnte.“

Wildauer ist selbst Molekularbiologe und hat sich erst im Zuge seiner beruflichen Tätigkeit der Datenanalyse zugewandt. Die beiden Data Scientists, die in Wildauers Team arbeiten, bringen hingegen keinen Hintergrund aus den Life Sciences mit, sondern kommen aus technischen Fächern, in denen sie sich intensiv mit statistischen Methoden beschäftigt haben. „Natürlich ist es von Vorteil, wenn man ein Grundverständnis mitbringt und weiß, was ein pH-Wert und was Chromatographie ist. Die Methoden, die wir anwenden, verlangen das aber nicht.“ Es gebe sogar Fälle, in denen die allzu große Vertrautheit mit dem Produktionsprozess eher hinderlich sei: „Ein routinierter Operator kann weniger unvoreingenommen an die Fehleranalyse herangehen“, sagt Wildauer.

Zur weiteren Verstärkung seines Teams sucht der „Head of Digital & Data Science“ Menschen, die einen soliden Hintergrund in statistischer Datenanalyse haben: „In den vergangenen Jahren sind mehrere Studiengänge auf diesem Gebiet entstanden. Wichtig ist aber nicht, dass die Studienrichtung „Data Science“ im Namen hat, sondern, dass die Absolventen gute Programmierkenntnisse in den für die Data Science essenziellen Sprachen Python und R erhalten“, so Wildauer.



Thomas Rattei, Professor für Computational Biology an der Uni Wien: „Die Bio Data Science wendet Methoden der Bioinformatik an, die Bioinformatik lässt sich von der Data Science inspirieren.“



Michael Wildauer, Head of Digital & Data Science, Takeda Wien: „Wichtig ist, dass die Absolventen gute Programmierkenntnisse in den für die Statistik essenziellen Sprachen Python und R erhalten.“



Milica Kronic, Studiengangsleiterin Bio Data Science: „Das Besondere an unserem Studiengang ist, dass die nötigen Programmierkenntnisse von Grund auf im Studium erlernt werden.“

► Wo die Disziplinen einander berühren

Thomas Rattei war an der Konzeption des Studiengangs „Bio Data Science“ am Tullner Standort der FH Wiener Neustadt – selbst institutionelles Mitglied der ÖGMBT – beteiligt, der genau diese Kompetenzen vermitteln will. „Das Besondere an unserem Studiengang ist, dass die nötigen Programmierkenntnisse von Grund auf im Studium erlernt werden“, sagt dazu Studiengangsleiterin Milica Kronic. Die Teilnehmer ler-

nen, Daten selbstständig zu analysieren und zu interpretieren und erhalten fundierte Kenntnisse der statistischen Werkzeuge, die man dafür benutzt. „Ich habe einige Studierende kennengelernt, die in unserer Gruppe ihre Master-Arbeiten gemacht haben. Die kamen aus der Biochemie oder der Medizin und wollten sich auf dem Gebiet der Datenanalyse vertiefen“, erzählt Rattei.

An den vordersten Fronten der wissenschaftlichen Entwicklung verschwimmen die scharfen Grenzen zwischen den Disziplinen mitunter. Eines der schwierigsten

Probleme der Bioinformation ist heute die Vorhersage der dreidimensionalen Struktur von Proteinen, wenn ihre Aminosäuresequenz gegeben ist. „Mechanistisch sind wir noch nicht in der Lage, zu sagen, wie die Kette sich falten wird“, sagt Rattei. In internationalen Projekten wie „AlphaFold“ versucht man daher erfolgreich, dem Problem mit „Deep Learning“-Verfahren zu begegnen. „Für Forschungsaufgaben wie dieser nimmt sich die Bioinformatik heute auch Anleihen bei dem, was aus der Data Science kommt“, so Rattei. (gs) ■

Wählen Sie die bessere Technologie Präzise, vielseitige Schlauchdosierpumpen

qdos
Peristaltic Metering

Fördermengen von 0,1 bis 2.000 ml/min bis 7 bar
Direkter Einbau in die Anlage ohne zusätzliches Zubehör
Werkzeugfreie Wartung in unter einer Minute



qdos