Data mining: new opportunities for medicine and public health

Research and healthcare activities produce huge quantities of data that need to be presented in an understandable structure. This requires computer-assisted extraction of relevant data and the use of statistical methods. This process, known as data mining, enables the discovery of patterns in large data sets. Data mining methods are of particular importance in fields that use high-throughput methods, visualisation methods and telemedical applications.

Experts from the fields of computer science, mathematics and statistics cooperate with specialists from the respective fields of application in order to select, process and analyse data according to specific criteria. Such specialists include biologists, medical, pharmaceutical and surgical researchers. Professions such as bioinformaticians and systems biologists are starting to appear and have begun exploiting the enormous potential of this development.

These days, people often share their workplaces with huge, computer-controlled equipment.

© GATC Biotech AG

Turning mounds of data into useful information for decision-making

Researchers and scientists in the fields of medicine and research have to deal with data volumes in the terabyte and petabyte range. It is estimated that the mounds of data currently generated worldwide double every two years. The term 'big data' is an attempt to describe the extent of the data avalanche that is engulfing scientists and institutions. Rapid developments in genome and proteome research are dramatically changing the life sciences. The rapid progress in genome sequencing produces ever growing amounts of data. However, these data only become useful if they can be effectively organized and if data mining algorithms for determining and interpreting interesting patterns are available.

Practical use in the healthcare sector

Laboratory assistant analyzing data on a computer.

© GATC Biotech AG

Data mining can be of economic, medical as well as organizational value for all parties involved in the healthcare sector. Data resources in hospitals and clinics can be used as a basis for financial and internal controls, medical process management and risk identification. Various data mining methods and processing models are able to transform data into useful information for professional interpretation and decision-making. Information obtained with data mining methods, based for example on comparative data from medical analyses, can also help physicians in the prevention and diagnosis of disease as well as in patient therapy and care.

Simulation of living cells

Baden-Württemberg is home to a large number of biotech companies with technological core competencies that enable them to focus on the management of big data. Two of these companies, Insilico Biotechnology AG in Stuttgart and quantiom bioinformatics in Weingarten, are presented below in greater detail.

Bioinformatics and protein design go hand in hand.
© BIOPRO/Bächtle

Insilico Biotechnology AG from Stuttgart designs and optimizes biotechnological processes for the chemical, pharmaceutical and food industries. "We provide software solutions for simulating the behaviour of cells or organisms. This knowledge enables us to reduce the time required for the development or optimisation of biotechnological processes involving the production of drugs, for example," said Klaus Mauch, CEO of Insilico. The company owns a worldwide unique systems biology platform that integrates proprietary databases, cell models and computer-assisted analysis methods. Insilico offers new solutions based on the integration and analysis of experimental data using genome-wide network models for the production of biochemicals and biopharmaceuticals as well as for the validation of drug candidates at an early stage. Customers include big industrial companies such as Boehringer Ingelheim Pharma GmbH & Co. KG headquartered in the Baden-Württemberg city of Biberach.

Innovative patient monitoring

The company quantiom bioinformatics from Weingarten has developed a highly effective data mining software that is successfully used in hospitals and for research. The software facilitates the reading of measurement profiles of individual patients. The company's Generic Signal Profiler enables the automated examination of patient profiles and is able, amongst other things, to recognize whether a patient is suffering from sleep-disordered breathing and respiratory conditions. According to quantiom, the RNA Integrity Number (RIN) software tool, which was developed using the company's Generic Signal Profiler, has already become the industry standard for estimating the integrity of RNA (ribonucleic acid) in thousands of laboratories worldwide.

Automated image analyses

A revolutionary new method for the automated tracking of the movement of biological particles in cell microscopy images achieved the best overall result in an international competition in Barcelona in 2012 that compared different image analysis methods. The probabilistic particle tracking method was developed by Dr. William J. Godinez and Dr. Karl Rohr in close cooperation with researchers from the University Hospital in Heidelberg. The method is used for studying infectious diseases caused by hepatitis C viruses.

Methods with huge future potential

Data mining techniques that can manage thousands of data sets on standard computers are already available today. Additional data mining potential arises from the increasing storage capacity of computers and the progressive integration of databases in the fields of medicine, health and research.

Heike Laue - 28.04.2014
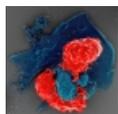© BIOPRO Baden-Württemberg GmbH

Dossier

28-Apr-2014

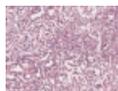## Other dossier articles

04.05.2026
Fractures in focus: predicting rather than waiting

12.02.2026
Intelligent immunotherapy – safety ensured

25.09.2025
Powerful AI systems using synthetic training data