

2. BIOSENSOR SYMPOSIUM

TÜBINGEN 2001

<http://barolo.ipc.uni-tuebingen.de/biosensor2001>

Data Mining trifft Bioinformatik: automatisierte Interpretation komplexer genomischer Zusammenhänge in Verbindung mit der DNA-Chip-Technologie

W. Dubitzky, S. Bulashevskaya, D. Berrar, C. Conrad, M. Granzow und R. Eils

Abteilung "Intelligente Bioinformatiksysteme", Deutsches Krebsforschungszentrum, 69120 Heidelberg, Deutschland

r.eils@dkfz-heidelberg.de, <http://www.iwr.uni-heidelberg.de/groups/bioinf/>

Tutorial

Die letzten Jahre haben einen dramatischen Anstieg im Umfang der genetischen Information, die im elektronischen Format gespeichert wird, gezeigt. Die Informationsmenge in den Bereichen Genomics und Proteomics verdoppelt sich schätzungsweise alle 20 Monate, und Umfang und Anzahl der Datensätze nehmen sogar noch schneller zu. Einhellige Meinung besteht darüber, dass eine hochentwickelte Auswertung solcher Daten in unterschiedlichen Bereichen wie beispielsweise der Genetik von Krankheiten und der Pharmacogenomics-Forschung von größter Bedeutung ist. Während sich sowohl Bioinformatikfirmen als auch institutionelle Einrichtungen darauf konzentriert haben, heterogene Daten in Genomics und Proteomics zu integrieren, steht eine systematische Datenauswertung funktioneller genomischer und proteomischer Daten noch im Anfangsstadium. Obwohl Data-Mining in finanz- und betriebswirtschaftlichen Gebieten wie im Einzelhandel und Finanzwesen viele Erfolge verzeichnen konnte, ist seine Anwendung auf Daten in Wissenschaft und Technik nicht ohne weiteres möglich. Die Gründe dafür sind vielfältig:

Datenmengen sind in den Biowissenschaften häufig beträchtlich größer im Volumen, strukturell komplexer als traditionelle Businessdaten und weisen hohe temporäre Änderungsraten auf.

Im Gegensatz zum Businessmilieu ist in den Biowissenschaften der Umfang des bestehenden Hintergrundwissens ungleich größer.

Wir werden über unsere aktuellen Bemühungen berichten, die Technologie des Data Mining auf eine Anpassung in den Biowissenschaften anzupassen. Insbesondere wird gezeigt, wie wissensbasiertes Data-Mining in Kombination mit der DNA-Chiptechnologie für eine systematische Aufdeckung komplexer funktioneller Beziehungen zwischen genotypischer Information und klinischen Parametern eingesetzt werden kann.