Naturwissenschaft

Frank Thierolf

Mathematische Modelle und Methoden zur Genexpressionsanalyse in der Bioinformatik

Diplomarbeit

Diplom.de

Bibliografische Information der Deutschen Nationalbibliothek:

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de/ abrufbar.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlages. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Copyright © 2003 Diplomica Verlag GmbH
ISBN: 9783832472139

http://www.diplom.de/e-book/222517/mathematische-modelle-und-methoden-zur-genexpressionsanalyse-in-der-bioinformatik

Frank Thierolf

Mathematische Modelle und Methoden zur Genexpressionsanalyse in der Bioinformatik

Diplom.de

Frank Thierolf

Mathematische Modelle und Methoden zur Genexpressionsanalyse in der Bioinformatik

Diplomarbeit
Technische Universität Darmstadt
Fachbereich Mathematik
Abgabe Juli 2003

Diplom.de

Diplomica GmbH
Hermannstal 119k
22119 Hamburg

Fon: 040 / 655 99 20
Fax: 040 / 655 99 222

agentur@diplom.de
www.diplom.de

ID 7213

ID 7213
Thierolf, Frank: Mathematische Modelle und Methoden zur Genexpressionsanalyse in der Bioinformatik
Hamburg: Diplomica GmbH, 2003
Zugl.: Technische Universität Darmstadt, Technische Universität, Diplomarbeit, 2003

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtes.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Die Informationen in diesem Werk wurden mit Sorgfalt erarbeitet. Dennoch können Fehler nicht vollständig ausgeschlossen werden, und die Diplomarbeiten Agentur, die Autoren oder Übersetzer übernehmen keine juristische Verantwortung oder irgendeine Haftung für evtl. verbliebene fehlerhafte Angaben und deren Folgen.

Diplomica GmbH
http://www.diplom.de, Hamburg 2003
Printed in Germany

Inhalt

Vorwort **4**

1 Einleitung und Überblick **6**

2 Biologische Grundlagen **8**
2.1 Proteine 8
2.2 DNA 10
2.3 RNA 12
2.4 Genexpression 12
2.4.1 Transkription 13
2.4.2 RNA-Prozessierung 13
2.4.3 Translation 14
2.4.4 Genregulation 15
2.5 DNA-Mikroarrays 15

3 Boolesche Netzwerke als Modelle der Genregulation **18**
3.1 Das binäre Modell 19
3.2 Identifikation boolescher Netzwerke 22
3.2.1 Ein einfacher Algorithmus: BOOL-1 23
3.2.2 Laufzeitbetrachtung für BOOL-1 24
3.2.3 Informationsbedarf 24
3.2.4 Behandlung von Störungen 26
3.3 Der Reverse-Engineering-Algorithm 27
3.3.1 Quantifizierung von Information 27
3.3.2 REVEAL - der eigentliche Algorithmus 30

4 Clustering-Methoden **35**
4.1 Euklidische Cluster-Analyse 36
4.1.1 Konstruktion eines Euklidischen Distanzbaumes mittels des Unweighted-Pair-Grouping-Method-Algorithm 38
4.1.2 Nicht-hierarchisches Clustering mit K-means 39
4.2 Alternative Ansätze zur Cluster-Analyse 44

5 Bayesianische Netzwerke als Modelle der Genregulation **47**
5.1 Grundbegriffe der Stochastik 47
5.2 Grundlagen Bayesianischer Netzwerke 50
5.3 Scoring von Graphen 52
5.4 Nicht-parametrischer Regressionsansatz 54

6 Differentialgleichungssysteme als Modelle der Genregulation **57**
6.1 Die Grundidee 57
6.2 Ansatz von Iba und Sakamoto 58
6.2.1 Genetische Programmierung 58
6.2.2 Methode der kleinsten mittleren Quadrate 60
6.2.3 Der Algorithmus 65

7 Diskussion **68**
7.1 Modellvergleich 68
7.2 Ausblick 71

Abkürzungsverzeichnis **73**

Abbildungsverzeichnis **75**

Literaturverzeichnis **78**

Index **82**

Vorwort

Kaum ein Forschungsgebiet steht derzeit so sehr im Zentrum der öffentlichen Beachtung wie die noch relativ junge Genomforschung. Für Aufsehen sorgte zuletzt die nun fast vollständige Entschlüsselung des menschlichen Erbguts. Spekulationen über die sich damit eröffnenden Möglichkeiten - ob sehr wünschenswert oder teils ethisch fragwürdig - scheinen keine Grenzen gesetzt.

Tatsächlich werden sicher noch Jahre oder sogar Jahrzehnte vergehen, bis sich die durch die Entschlüsselung des Erbguts bietenden Möglichkeiten wirklich nutzen lassen werden. Ist es mittels der Gentechnologie nunmehr zwar gelungen, die in den rund fünfzigtausend menschlichen Genen gespeicherten Informationen nahezu vollständig auszulesen und elektronisch zu speichern, so war dies erst der Anfang. Fragen nach Zusammenhängen wie etwa zwischen Beschaffenheit des Erbguts und Krankheiten wie Alzheimer oder Krebs lassen sich bislang noch nicht oder nur ungenügend beantworten.

Die Entschlüsselung des menschlichen Erbguts war sicher eine große Herausforderung für die Wissenschaft und unabdingbare Voraussetzung für alle nun folgenden Schritte. Letztere sind jedoch genauso anspruchsvoll und mit großer Wahrscheinlichkeit sogar noch bedeutender als der zuletzt erreichte Meilenstein der Genomforschung.

So richtet sich das Augenmerk nun immer mehr auf die Identifikation bestimmter Abschnitte des menschlichen Genoms, die z.B. mit Krankheiten in Zusammenhang stehen sowie auf die Analyse eben dieser Zusammenhänge.

Trotz des enormen medialen Interesses an der Genforschung wird nur selten deutlich, welche fachübergreifenden Anstrengungen sich hinter den veröffentlichten Ergebnissen verbergen. So dürfte sich nur den wenigsten Außenstehenden erschlossen haben, welche bedeutsame Rolle der Mathematik beim Erreichen des heutigen Kenntnisstandes zukam.