Statistische und numerische Methoden der Datenanalyse

Von Prof. Dr. rer. nat. Volker Blobel
und Prof. Dr. rer. nat. Erich Lohrmann
DESY und Universität Hamburg

B. G. Teubner Stuttgart · Leipzig 1998

Prof. Dr. rer. nat. Volker Blobel

Studium in Braunschweig und Hamburg mit Promotion 1968, wiss. Tätigkeit an der Universität Hamburg und am Deutschen Elektronen-Synchrotron DESY, seit 1977 Professor an der Universität Hamburg, zwischen 1979 und 1996 verschiedene Forschungsaufenthalte am CERN (European Laboratory for Particle Physics) in Genf.
Arbeitsgebiet: experimentelle Hochenergiephysik.

Prof. Dr. rer. nat. Erich Lohrmann

Studium in Stuttgart, Promotion 1956 mit einer Arbeit über Prozesse sehr hoher Energie in der kosmischen Strahlung, Arbeiten über kosmische Strahlung und Hochenergiephysik an den Universitäten Bern, Frankfurt und Chicago, seit 1961 am Deutschen Elektronen-Synchrotron, seit 1976 Professor an der Universität Hamburg, Mitglied des CERN Direktoriums 1976–78, Zuständigkeit Forschung und Datenverarbeitung.
Gegenwärtiges Arbeitsgebiet: experimentelle Hochenergiephysik.

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Blobel, Volker:
Statistische und numerische Methoden der Datenanalyse / von Volker Blobel ; Erich Lohrmann. – Stuttgart ; Leipzig : Teubner, 1998
(Teubner-Studienbücher : Physik)
ISBN 978-3-519-03243-4 **ISBN 978-3-663-05690-4 (eBook)**
DOI 10.1007/978-3-663-05690-4

Vorwort

Der Umfang des Datenmaterials in Wissenschaft und Technik nimmt immer schneller zu; seine Auswertung und Beurteilung erweisen sich zunehmend als die eigentliche Schwierigkeit bei vielen wichtigen Problemen. Dem steht auf der anderen Seite ein seit Jahren ungebrochenes Anwachsen der Rechnerleistung und die zunehmende Verfügbarkeit mächtiger Algorithmen gegenüber, aber es ist oft nicht einfach, von diesen Hilfsmitteln den richtigen und professionellen Gebrauch zu machen. Dieses Buch, entstanden aus der Praxis der Verarbeitung großer Datenmengen, will eine Einführung und Hilfe auf diesem Gebiet geben.

Viele der Probleme sind statistischer Natur. Hier ist es sprichwörtlich leicht, Fehler zu machen. Deshalb sind der Erklärung und der kritischen Durchleuchtung statistischer Zusammenhänge auch im Hinblick auf die Praxis ein angemessener Raum gewidmet und ebenso den Monte Carlo-Methoden, welche heute einen verhältnismäßig einfachen Zugang zu vielen statistischen Problemen bieten.

Werkzeuge für die Organisation und Strukturierung großer Datenmengen bilden ein weiteres wichtiges Thema. Dazu gehören auch effiziente Verfahren zum Sortieren und Suchen, welche oft Teil größerer Algorithmen sind.

Die Verarbeitung großer Datenmengen hat oft die Extraktion verhältnismäßig weniger Parameter zum Ziel. Hier sind Verfahren wie die Methoden der kleinsten Quadrate und der Maximum-Likelihood wichtig, in Verbindung mit effektiven Optimierungsalgorithmen. Ein weiteres Problem, welches oft unterschätzt wird, ist die Rekonstruktion ursprünglicher Verteilungen aus fehlerbehafteten Messungen durch Entfaltung.

Mit der Verfügbarkeit mathematischer Bibliotheken für Matrixoperationen können viele Probleme elegant in Matrixschreibweise formuliert und auf Rechnern gelöst werden. Deswegen werden auch die einfachen Grundlagen der Matrixalgebra behandelt.

Das Buch führt in diese Dinge ein, es ist gedacht für Naturwissenschaftler und Ingenieure, die mit Problemen der Datenverarbeitung befaßt sind. Neben Algorithmen dienen zahlreiche einfache Beispiele der Verdeutlichung.

Um der Unsitte entgegenzuwirken, mehr oder weniger gut verstandene Rezepte schwarze-schachtelartig anzuwenden, wird auf Erklärungen und Beweise innerhalb eines vernünftigen Rahmens großer Wert gelegt, ebenso auf eine kurze Einführung des Informationsbegriffs.

Programmcodes von im Text erwähnten Algorithmen sind, sofern sie einen solchen Umfang haben, daß Kopieren von Hand nicht ratsam ist, im Internet zugänglich gemacht (`http://www.desy.de/~blobel`). Sie sind von den Autoren benutzt und überprüft worden, trotzdem kann für ihre Korrektheit und für die Abwesenheit von Nebeneffekten keine Garantie übernommen werden. Kurze Programme sind meist in Fortran, in einigen Fällen in C formuliert. Sie sollen

zur Präzisierung der Algorithmen dienen. Sie sind kurz genug, so daß sie jeder in seine Lieblingssprache umschreiben kann.

Vielen Kollegen, vor allem Dr. F. Gutbrod und Dr. D. Haidt sind wir dankbar für wertvolle Diskussionen, Informationen und Anregungen. Frau A. Blobel und Frau U. Rehder danken wir für ihre Hilfe bei der Erstellung des Manuskripts, Herrn Dr. P. Spuhler vom Teubner-Verlag danken wir für die gute Zusammenarbeit.

Hamburg, im Juli 1998 V. Blobel und E. Lohrmann

Inhaltsverzeichnis