

Jahresbericht der Arbeitsgruppe
Prof. Dr. U. Faigle
Prof. Dr. R. Schrader

**Zentrum für Angewandte Informatik
an der Universität zu Köln**



1999/2000

Impressum:

Herausgeber: Prof. Dr. U. Faigle, Prof. Dr. R. Schrader
Anschrift: Zentrum für Angewandte Informatik Köln (ZAIK)
Weyertal 80
50931 Köln
Telefon: 0221 / 470-6030
Fax: 0221 / 470-5160
Redaktion: Ch. Hagemeier, M. Peter, D. Rübiger, A. Schliep
Layout: Ch. Gawron, Ch. Hagemeier, A. Schliep
Umschlag: R. Boening, A. Schliep

Inhaltsverzeichnis

Vorwort	5
Überblick	6
1 Kombinatorische Optimierung	8
1.1 Generierung saisonaler Charterflugpläne	8
1.2 Fahrereinsatzplanung im ÖPNV	10
1.3 Tourenplanung	10
1.4 Steinerprobleme	12
1.5 Ein Färbungsproblem aus der Automobilproduktion	12
1.6 Testmengen in der ganzzahligen Optimierung	12
2 Forschungsschwerpunkt Verkehr	14
2.1 Wissenschaftlicher Fortschritt	14
2.2 Anwendungsprojekte	17
2.3 Simulation von Bahnanlagen	22
3 Bioinformatik	24
3.1 Ein neues Verfahren zum Lernen von Hidden-Markov-Modellen	24
3.2 Analyse von Proteinsequenzen für die Strukturvorhersage von Proteinen	25
3.3 Auswahl spezifischer Proben für DNA Arrays	27
3.4 Klassifizierung von NMR-Spektren	29
3.5 Effizientes Screenen von Klon-Bibliotheken	29
3.6 Modellierung metabolischer Netzwerke	31
3.7 Bioinformatik Interessengruppe BIG	32
4 Umweltmanagement	33
4.1 Modellierung und Simulation von Emissions-Minderungsaktivitäten am Beispiel von Joint-Implementation unter Verwendung des TEM-Modells	33
4.2 Grundgleichungen des TEM-Modells	34
4.3 Simulation	34
4.4 Umweltlizenzen	35
4.5 Umweltsystemforschung	35
5 Multimedia	36
5.1 Trial Solution	36
5.2 Graph Animation Toolbox	37
5.3 CATBox	39

5.4	Nikos	39
6	Modellierung und Simulation von Bausparkollektiven	42
6.1	Ein deterministisches Modell für Bausparkollektive	42
6.2	Weiterentwicklung des Modells	44
6.3	SVD-Clusterung von Bauspar-Zeitreihen	46
6.4	Stochastische Modellierung mit Hidden–Markov–Modellen	46
6.5	Ausblick	49
A	Computer und Netzwerke	50
A.1	Zentrale Server	50
A.2	Arbeitsplatzrechner	50
A.3	Vernetzung	50
B	Gäste	52
C	Veranstaltungen	54
D	Arbeitsergebnisse	56
E	Mitarbeiter 1999/2000	62

Vorwort

Das im Jahr 1997 durch die Zusammenlegung des Universitätsrechenzentrums und des Zentrums für Paralleles Rechnen entstandene Zentrum für Angewandte Informatik der Universität zu Köln (ZAIK) hat sich in den letzten Jahren zu einer zentralen wissenschaftlichen Einrichtung entwickelt, der zehn Arbeitsgruppen aus den Bereichen Informatik, Angewandte Mathematik, Wirtschaftsinformatik, medizinische Informatik und sprachliche Informationsverarbeitung angehören. Dadurch und durch die im Jahr 1999 erfolgte Wiederbesetzung der Professur von Herrn Bachem stellt sich die Arbeitsgruppe zum ersten Mal mit einem Jahresbericht als gemeinsame Arbeitsgruppe Faigle/Schrader innerhalb des ZAIK vor.

Durch die Wiederbesetzung ist der Schwerpunkt unserer Arbeiten – die Weiterentwicklung der theoretischen Grundlagen der algorithmischen Mathematik und ihre Umsetzung in praxisnahe Anwendungen – gestärkt worden. Wir hoffen, mit dem vorliegenden Bericht deutlich machen zu können, dass unsere Arbeiten auf den Tätigkeitsfeldern Kombinatorische Optimierung, Produktions- und Logistikplanung, Verkehrssimulation und die Zusammenarbeit mit den Landesbausparkassen erfolgreich fortgesetzt werden konnten. Neu hinzugekommen, z. T. durch den Wechsel in der personellen Zusammensetzung bedingt, sind die Felder Bioinformatik und Umweltmanagement.



Rainer Schrader



Ulrich Faigle

Unser Dank gilt allen, die uns wie in der Vergangenheit materiell und ideell unterstützt haben. Dazu gehören die Deutsche Forschungsgemeinschaft, die Europäische Union, das Deutsche Akademische Auslandsamt, das Bundesministerium für Wissenschaft und Forschung, das Ministerium für Schule, Wissenschaft und Forschung, das Zentrum für Luft- und Raumfahrt, die Studienstiftung des Deutschen Volkes, die Alfred-Krupp-Stiftung, der DFN-Verein sowie Firmen und Verbände. Das Rektorat, viele Kollegen in Köln, der Kanzler und seine Verwaltung haben uns – insbesondere in der Gründungsphase des ZAIK – mit Rat und Tat unterstützt.

Besonderer Dank gilt allen Mitarbeiterinnen und Mitarbeitern für Ihren Einsatz, ihre kritischen Anregungen und den Enthusiasmus für die gemeinsame Sache.

Ulrich Faigle R. Schrader

Überblick

Auch in diesem Berichtszeitraum bildete die Kombinatorische Optimierung und ihre Anwendungen einen Schwerpunkt unserer Arbeit. Langjährige Projekte im Bereich der Flugplanerstellung für Luftverkehrsgesellschaften wurden kontinuierlich weiter verfolgt. Zusätzlich lieferte eine Studie zur rechnergestützten Fahrereinsatzplanung im ÖPNV gute Ergebnisse.

Im Bereich der Tourenplanung wurden neue Verfahren zur Lösung von Routingproblemen mit Umlademöglichkeit entwickelt und erfolgreich eingesetzt. Bei der Lösung des kapazitierten Routingproblems gelang ein Weltrekord, indem erstmals zwei schwierige Benchmark-Probleme in angemessener Zeit exakt gelöst wurden.

Außerhalb von Projekten runden im Bereich der kombinatorischen Optimierung Arbeiten an Färbungsproblemen und Testmengen die theoretische Forschung ab.

Die Arbeiten aus dem Bereich der Verkehrssimulation waren geprägt durch interdisziplinäre Kooperationen, in denen die vorhandenen Modelle insbesondere unter Anwendungsaspekten weiterentwickelt wurden. So werden von uns im Rahmen eines Sonderforschungsbereiches Umweltwirkungen des Straßenverkehrs untersucht und quantifiziert. Um ein tieferes Verständnis von Mobilität zu erhalten und mögliche Lenkungsinstrumente zu evaluieren, entwickeln wir in Zusammenarbeit mehrerer Partner ein Simulationswerkzeug, in dem unsere Modelle mit Modellen der Verkehrsnachfrage gekoppelt werden. Im Rahmen des Gigabit Testbed West haben wir untersucht, wie große Straßennetze verteilt simuliert und visualisiert werden können.

In der Bioinformatik wurde die erfolgreiche Zusammenarbeit — u.a. mit dem Institut für Biochemie an der Universität Köln im Bereich der Erkennung von entfernt homologen Proteinsequenzen und dem Los Alamos National Laboratory bei der Entwicklung effizienter Verfahren zur Erstellung physikalischer Kartierungen — ausgebaut und eine stärkere Vernetzung der Aktivitäten erreicht. Gemeinsam mit dem Institut für Genetik wurde ein Algorithmus

zum Design sogenannter DNA-Chips entwickelt. Dieser erlaubt es, auch bei Anwendungen mit einer großen Anzahl an Zielsequenzen bzw. –organismen, hochspezifische Proben effizient zu berechnen. Weiterhin wurde ein neues Verfahren zur statistischen Modellierung mit Hidden-Markov-Modellen (HMM) entwickelt. Statt der bisher zur Analyse biologischer Sequenzen eingesetzten sogenannten Profile-HMM können mit diesem statistisch robusten Werkzeug allgemeine HMM als Black-Box-Modelle Verwendung finden.

Seit der Konferenz 1992 in Rio de Janeiro, auf dem die erste Klimaschutzrahmenkonvention verabschiedet wurde, sehen es die unterzeichnenden Staaten als gemeinsame Aufgabe an, die Klimaproblematik wissenschaftlich zu untersuchen und den möglichen Umweltveränderungen zu begegnen. Auf der 1. Vertragsstaatenkonferenz 1995 zum Rahmenübereinkommen über Klimaveränderungen in Berlin wurde daraufhin beschlossen, Joint-Implementation Programme in einem verstärkten Maße einzusetzen. Joint-Implementation meint hierbei die gemeinsame Umsetzung von Klimaschutzziele im Rahmen von Technologiek Kooperationen. Da bisher nur geringe Projekterfahrung mit Joint-Implementation vorliegt, ist man auf eine Simulation von verschiedenen Handlungsszenarien angewiesen. Um eine solche Analyse auf komfortable Weise zu ermöglichen, wurde das TEM-Modell (Technologie-Emissionen-Mittel-Modell) entwickelt und analytisch untersucht. Die Optimierung von Investitionsausgaben mithilfe des TEM-Modells trägt auf diese Weise zu einem Umweltmanagement bei, in dessen Zentrum die Erfüllung des Kyoto-Protokolls steht.

Das ZAIK ist in mehreren Projekten an der Gestaltung und Weiterentwicklung von multimedialen Lehr- und Lerneinheiten beteiligt. Im Rahmen des europäischen Projektes "Trial-Solution" wird die sogenannte Slicing-Book-Technologie entwickelt und erprobt, die es ermöglicht,

Lehr- und Lernmaterialien individuell zu modularisieren und multimedial aufzubereiten. Zu den ausgewählten Materialien wird die Combinatorial Algorithm Toolbox (CATBox), die 2001 im Springer Verlag erscheinen wird, zählen. Bei ihr handelt es sich um einen interaktiven Kurs in diskreter Mathematik, der am ZAIK entwickelt wurde. CATBox basiert auf der Graphenvisualisierungssoftware Gato (Graph Animation Toolbox). Im Mittelpunkt des NIKOS-Projektes, das gemeinsam mit dem Seminar für Politikwissenschaft durchgeführt wurde, stand die Entwicklung eines virtuellen ONLINE-Seminars für Studierende aus dem beteiligten Fachgebiet. Die computergestützte interaktive Simulation von politischen Handlungszusammenhängen in einem virtuellen Entscheidungsraum konnte insbesondere das selbstorganisierte Lernen von komplexen Entscheidungsprozessen

besonders fördern.

Innerhalb der langjährigen Zusammenarbeit zwischen dem ZAIK und den Landesbausparkassen wurde ein deterministisches Modell zur Simulation von Bausparkollektiven entwickelt, welches auf der Clusteranalyse realer Bausparverträge basiert. Im Laufe der letzten Jahre wurde dieses Modell im Hinblick auf seinen praktischen Einsatz für die Landesbausparkassen stetig verfeinert und verbessert mit dem Ziel, eine flexiblere und effizientere Handhabung des Modells zu erreichen. Daneben wurde im Rahmen von Promotionen ein neuer Modellansatz untersucht, bei dem von einer deterministischen hin zu einer stochastischen Modellierung von Bausparkollektiven mit Hilfe von Hidden-Markov-Modellen übergegangen wurde.

Kapitel 1

Kombinatorische Optimierung

1.1 Generierung saisonaler Charterflugpläne

Durch die Globalisierung der Märkte und Internationalisierung von Unternehmen wird der effiziente Einsatz personeller und materieller Ressourcen zu einem wesentlichen Faktor für die Konkurrenzfähigkeit von Transport- und Logistikunternehmen. Im Bereich von Linienfluggesellschaften wurden bereits in den letzten Jahren Optimierungsverfahren für zwei Teilproblemstellungen in diesem Bereich am ZAIK entwickelt: sowohl ein Branch and Price Ansatz für das Fleet-Assignment-Problem als auch ein auf diesem Ansatz basierendes Lösungsverfahren für das Crew-Scheduling-Problem.

Die entsprechenden Planungsprobleme unterscheiden sich bei Charterfluggesellschaften in einigen zentralen Punkten. Insbesondere dem Problem der starken Saisonabhängigkeit und dem damit stärkeren Einfluss des Passagieraufkommens auf die Flugplangestaltung muss besondere Aufmerksamkeit geschenkt werden. Mit einem Partner aus dem Charterflugbereich haben wir uns der Lösung dieses Problems zugewandt.

1.1.1 Problemstellung

Eingabe für dieses Optimierungsproblem sind saison- und wochentagabhängige prognostizierte Nachfragen für alle angebotenen Flugstrecken. Diese bestehen aus den durch die enge Kopplung der Charterfluggesellschaft mit den Reiseveranstaltern garantierten Passagiernachfragen und den Daten der vergangenen Jahre. Ausserdem sind für alle Flugstrecken und Flugzeugtypen Kostensätze sowie entsprechende Erträge beförderter Passagiere gegeben.

Das Problem besteht nun darin, die zur Verfügung stehenden Flugzeuge der Charterfluggesellschaft so einzuplanen, dass ein unter Profitabilitäts Gesichtspunkten optimaler Flugplan entsteht. Der Gewinn eines Flugplanes

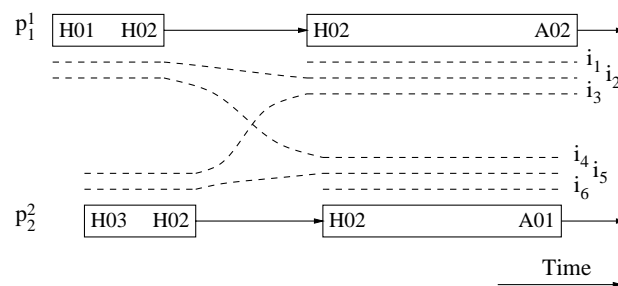


Abbildung 1.1: Verknüpfung der Rotationen mit Reiserouten

errechnet sich dabei sowohl aus den Fix- und Betriebskosten der Flugzeuge (wobei nicht eingesetzte Flugzeuge an andere Fluggesellschaften vermietet oder Flugzeuge angemietet werden können) als auch aus den oben erwähnten passagierabhängigen Kosten und Erträgen.

Weiter müssen operationelle Nebenbedingungen wie Flughafenöffnungszeiten, Mindestbodenzeiten (zum Auftanken und Catering) und Flugverbote auf bestimmten Flugstrecken beachtet werden.

1.1.2 Mathematische Modellierung und Lösungsansatz

Wir verwenden zur Lösung eine Set-Partitioning-Formulierung mit Nebenbedingungen, die als Lineares Programm modelliert wird: aus der Menge aller möglichen Tagesflugzeugrotationen (erzeugt durch ein Netzwerk, dessen Aufbau in Abbildung ?? verdeutlicht wird) muss eine Teilmenge ausgewählt werden, so dass, unter Beachtung von globalen Nebenbedingungen wie der Anzahl an verfügbaren Flugzeugen, möglichst viele Passagiere auf den gewünschten Flugstrecken befördert werden.

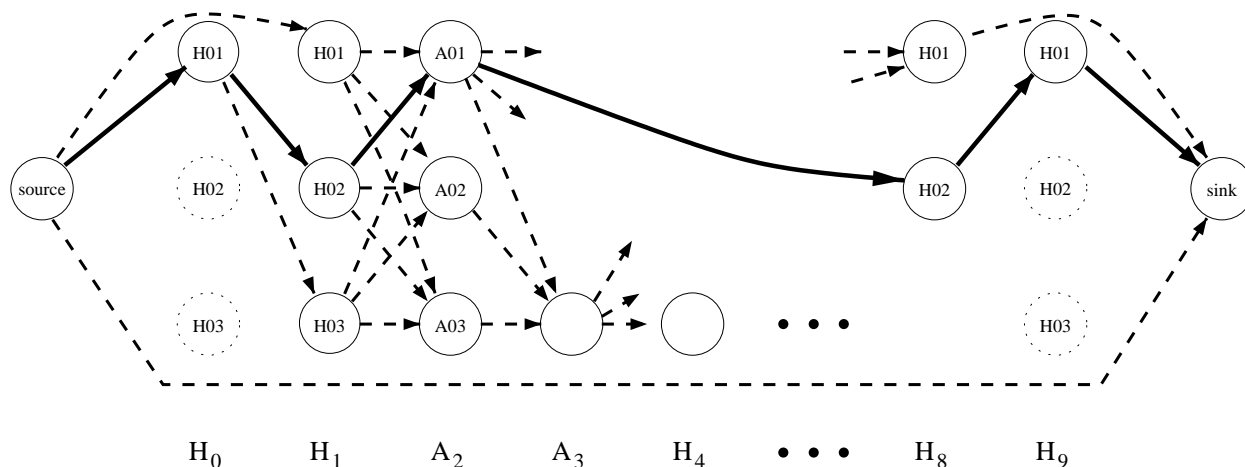


Abbildung 1.2: Aufbau des Netzwerks zum Column Generation

Da die Anzahl aller möglichen Flugzeugrotationen in einer handhabbaren Größenordnung liegt, können während des gesamten Lösungsprozesses auch bei größeren praxisbezogenen Datensätzen alle möglichen Tagesrotationen gleichzeitig beachtet werden (siehe unten).

Ein Problem stellt jedoch die Qualität der erzeugten Lösungen dar: Ohne zusätzliche Beschränkungen benötigt das Lösungsverfahren extrem lange, um akzeptable Lösungen zu erzeugen. Aus diesem Grunde werden sogenannte „Min-Cover“-Schnitte im Verlauf des Lösungsprozesses zum Linearen Programm hinzugefügt, damit die Wahl der möglichen Lösungen schnell eingeschränkt werden kann und so gute Lösungen in kurzer Zeit erzeugt werden können. In diesen Branch and Cut Ansatz konnten zusätzlich auch Heuristiken zum schnellen Auffinden zulässiger Lösungen integriert werden.

1.1.3 Ergebnisse

Uns stehen verschiedene Datensätze einer großen europäischen Charterfluggesellschaft zur Verfügung, die etwa 20 deutsche Flughäfen und 60 Zielflughäfen bei ca. 60 planbaren Flugzeugen enthalten. Die entsprechende Nachfrage an Flugverbindungen ist nach einzelnen Wochentagen aufgeschlüsselt, so dass in einer einfachen Version jeder Wochentag einzeln optimiert werden kann. Hieraus ergeben sich ungefähr 13.000 mögliche Tagesrotationen pro Wochentag, wenn nur direkte Flüge ohne Umsteigen oder Zwischenlandungen möglich sind. (Wie Abbildung ?? verdeutlicht, steigt die Komplexität des Problems an, wenn auch Umsteigen oder Zwischenlandungen erlaubt

sind.) Für die Optimierung über die ganze Woche ergeben sich knapp 60.000 mögliche Tagesrotationen, unter denen die beste Kombination für die gesamte Woche gesucht ist.

Bei der Generierung von Flugplänen für einzelne Wochentage erhalten wir Laufzeiten, die auf einem handelsüblichen PC im Bereich von Minuten liegen. Die so erzeugten Lösungen liegen meist höchstens 0,2% vom Optimalwert entfernt. Wird ein Wochenflugplan angestrebt und gleichzeitig die Möglichkeit eröffnet, dass Passagiere einmal das Flugzeug wechseln können, sowie pro Flugzeug eine Zwischenlandung pro Passagierreisroute eingeplant werden darf, so vergrößert sich die Laufzeit: Sie beträgt dann für eine Lösung, die höchstens 3–4% vom Optimalwert entfernt ist, wenige Stunden.

1.1.4 Aktuelle Forschung

Zur Zeit wird dieser Lösungsansatz in zwei Richtungen weiterentwickelt: An großen Flughäfen wird die hohe Nachfrage nach Starts und Landungen durch die Vergabe von sogenannten „Slots“ geregelt. Fluggesellschaften besitzen solche Zeitfenster für ihre Flugbewegungen an den entsprechenden Flughäfen; zu anderen Zeitpunkten sind keine Flugbewegungen an dem Flughafen planbar. Da Charterfluggesellschaften meist kleine (regionale) Flughäfen anfliegen, existieren nur auf wenigen Flughäfen Slots. Es werden zur Zeit verschiedene Möglichkeiten zur Berücksichtigung solcher Slots in unserem Branch and Cut Rahmen erprobt.

Die zweite Forschungsrichtung befasst sich mit der Integration von Wartungsereignissen in den zu erstellenden

Flugplan. Unter einem Wartungsereignis wird der Zeitraum verstanden, den die regelmäßige Wartung eines bestimmten Flugzeugtyps in Anspruch nimmt. Es ist nun dafür zu sorgen, dass alle Flugzeuge eines entsprechenden Typs in vordefinierten Abständen gewartet werden. Eine Modellierung und ein darauf basierender Lösungsansatz werden für dieses Problem zur Zeit am ZAIK entwickelt.

Kontakt: airlines@zpr.uni-koeln.de

1.2 Fahrereinsatzplan im ÖPNV

Seit am 1. Januar 1996 eine EU-Richtlinie von Bund und Ländern in deutsches Recht umgesetzt wurde, müssen die Kommunen die Nahverkehrsleistungen bei den Verkehrsunternehmen einkaufen. In diesem Zusammenhang stellt, bei fortschreitender Privatisierung und zunehmendem Wettbewerb auf dem ÖPNV-Markt, der effiziente Ressourceneinsatz einen immer wichtigeren Faktor dar.

1.2.1 Problemstellung

Es sollen Arbeitspläne für Fahrer erstellt werden, so dass jedes Fahrzeug auf jeder Linie immer einen Fahrer hat. Bei einer vorgegebenen Anzahl an Fahrern sollen möglichst sozialverträgliche Dienstpläne erzeugt werden. Dies kann sich beispielsweise in einer ausgewogenen Dienstlänge oder einer angenehmen Pausenlänge ausdrücken.

Als Eingabe dient ein Fahrzeugumlaufplan. Aus ihm geht hervor, welches Fahrzeug sich zu welchem Zeitpunkt an welcher Haltestelle befindet. Jedes beliebige Teilstück dieses Fahrzeugumlaufplans muss genau einem Fahrer zugeordnet werden.

Ausserdem müssen operationelle Nebenbedingungen beachtet werden. Diese bestehen auf der einen Seite aus vertraglichen und tarifrechtlichen Bedingungen (wie beispielsweise Pausenzeitregelungen und maximale Dienstlängen). Auf der anderen Seite müssen auch betriebliche Vorgaben, wie das Verhältnis von Teilzeit- zu Vollzeitarbeit, eingehalten werden.

1.2.2 Mathematische Modellierung und Lösungsansatz

Als Modell für dieses Problem kommt ein Lineares Programm als Set-Partitioning-Formulierung mit Nebenbedingungen zum Einsatz. Hierbei wird aus der Menge aller möglichen Dienste für die Fahrer die Kombination

gewählt, die zum einen alle Dienstelemente überdeckt und zum anderen alle Nebenbedingungen einhält, wobei möglichst geringe Kosten verursacht werden.

Problematisch hierbei ist jedoch die schon bei einer geringen Anzahl an Dienstelementen sehr große Anzahl an möglichen Diensten. Daher wird zur Lösung ein Branch and Price Ansatz verwendet, wobei als Subproblem zur Erzeugung der benötigten Dienste ein Kürzeste-Wege-Problem gelöst wird. Einige Heuristiken zur Beschleunigung dieser Suche wurden bereits erprobt und erfolgreich implementiert. Allein durch den Einsatz alternativer Labeleliminationsverfahren konnte ein Geschwindigkeitszuwachs von bis zu 30% erzielt werden.

1.2.3 Ergebnisse

Wir hatten die Möglichkeit, das Lösungsverfahren mit Real-World-Daten eines norddeutschen Verkehrsbetriebes zu testen. Hier zeigte sich, dass die erzielten Ergebnisse von sehr guter Qualität sind und in Laufzeiten von etwa einer Stunde (auf einem handelsüblichen PC) produziert wurden. Bei einem Datensatz mit rund 760 Dienstelementen an sieben Ablösepunkten mit zwei möglichen Geschäftsstellen als Anfangspunkte der Dienste konnte bereits nach gut 40 Minuten eine Lösung produziert werden, die nachweisbar maximal 1,1% von der besten möglichen Lösung entfernt ist.

Die hier aufgezeigten Punkte prädestinieren das Verfahren für den planerischen Einsatz in der Praxis.

Kontakt: combopt@zpr.uni-koeln.de

1.3 Tourenplanung

Die Tourenoptimierung wird als eine der größten Erfolgsgeschichten des Operations Research bezeichnet (*G. Laporte*). Das starke Interesse an diesem Gebiet in Industrie und Forschung liegt zum einen am wirtschaftlichen Potenzial der Tourenoptimierung, zum anderen macht ihr Reichtum an Struktur sie zu einem faszinierendem Forschungsgebiet. Auch in den Jahren 1999/2000 konnte die Optimierungsgruppe am ZAIK wieder einige Facetten von Tourenplanungsproblemen bearbeiten.

Im Rahmen der Zusammenarbeit mit der Firma Profi.S wurden Praxisprojekte durchgeführt, die sich durch eine starke Vernetzung der Produktions- und Logistikplanung auszeichneten, während im eher theoretischen Bereich an der optimalen Lösung von Vehicle Routing Problemen gearbeitet wurde. Desweiteren haben wir Algorithmen für

intermodale Transportprobleme entwickelt. Diese Arbeiten haben sowohl theoretische als auch praktische Früchte getragen.

1.3.1 Integration von Produktions- und Transportlogistik

Bei der Büromöbelfirma *dyes* bestehen aufgrund der *Just In Time*-Fertigung zahlreiche wechselseitige Abhängigkeiten zwischen Produktion und Distribution.

Den Kunden wird bereits bei der Bestellung ein Liefertermin kalenderwochengenau zugesagt. Bei der tagesgenauen Planung müssen dann sowohl die Bedürfnisse der Produktion nach ausgeglichener Auslastung der Maschinen berücksichtigt werden, als auch die der ausliefernden Spedition (SML Logistik- und Distributionssysteme), da nach ganz Deutschland und auch ins benachbarte Ausland geliefert wird. Hier galt es einen Algorithmus zu finden, der allen Anforderungen gleichzeitig gerecht wird.

Wir haben hierfür einen Ansatz entwickelt, der zunächst eine Tourenplanung durchführt und später die Produktion mit Hilfe eines linearen Programms so auf die Wochtage verteilt, dass eine möglichst gleichmäßige Auslastung der diversen Produktionslinien erreicht wird.

1.3.2 Branch-and-Cut für das Vehicle Routing Problem (VRP)

Routingprobleme sind aus kombinatorischer Sicht „schwere“ Probleme. Das heißt, dass wir nicht hoffen können, einen Algorithmus zu finden, der solche Probleme in jedem Fall in annehmbarer Zeit löst. Dies hat zur Entwicklung einer Vielzahl von Heuristiken geführt, also von Algorithmen, die auf den jeweiligen Anwendungsfall abgestimmt sind und hier eine vernünftige (wenn auch mathematisch zumindest nicht beweisbar optimale) Lösung liefern. Dennoch ist die Optimierung von Vehicle Routing Problemen immer eine faszinierende Forschungsaufgabe geblieben. Um die Schwierigkeiten zu verstehen, die das VRP beinhaltet, sollte man sich vor Augen führen, dass heute Instanzen des Traveling Salesman Problem mit mehreren tausend Knoten mit Branch-and-Cut-Methoden gelöst werden können, während beim VRP schon Instanzen mit über 70 Knoten eine große Herausforderung darstellen.

Am ZAIK wurde zu diesem Thema von Ulrich Blasum eine Dissertation angefertigt. Zur Lösung von VRP-Instanzen mittels Branch-and-Cut ist das Aufinden von

geeigneten, verletzten Ungleichungen von zentraler Bedeutung. Im Rahmen der Dissertation wurden bisher bekannte Ungleichungstypen weiterentwickelt und neue gefunden. Ausserdem wurden Heuristiken und neue exakte Methoden zur Identifikation verletzter Ungleichungen entwickelt.

Dadurch gelang es erstmals zwei klassische Benchmarkinstanzen mit 76 Knoten auf einem einzelnen Prozessor zu lösen. Der im Rahmen der Arbeit entwickelte Code ist sehr effizient und findet die Lösungen innerhalb weniger Stunden, bisher verwendete Methoden benötigten mehrere Tage auf Parallelcomputern.

1.3.3 Transportprobleme mit Umladen

Das wachsende ökologische Bewusstsein, ebenso wie die Überlastung der Verkehrsinfrastruktur, haben das Interesse an intermodalen Strategien im Gütertransport beständig steigen lassen. So werden zur Entlastung der Infrastruktur und der Umwelt oft intermodale Logistikketten vorgeschlagen. Ein ähnliches Anliegen haben Speditionen, die ihre Transportaufträge (zumindest teilweise) über ein *Konsolidierungszentrum* abwickeln.

Im Bereich der strategischen Planung beschäftigt sich das *Operations Research* schon seit längerer Zeit mit solchen Problemen, hierbei werden jedoch vor allem Fragen der Platzierung von Konsolidierungszentren behandelt und das Transportaufkommen nur abgeschätzt. Die taktische und operationelle Planung der Transportketten stellt jedoch sehr viel höhere Anforderungen, da hier die aktuelle Auftragslage genau dargestellt werden muss. Daher wächst die Nachfrage nach Algorithmen, die solche Problemstellungen bearbeiten können.

Im kommerziellen Bereich werden Planungstools für die rechnergestützte Optimierung solcher Aufgaben jedoch noch kaum angeboten, da der Wechsel des Transportmediums eine wesentliche Erschwerung des Problems bedeutet. Wir haben solche Aufgabenstellungen aus verschiedenen Blickwinkeln untersucht und Algorithmen zu ihrer Lösung entwickelt.

Durch komplexitätstheoretische Untersuchungen können wir zeigen, dass schon stark relaxierte Probleme dieser Art schwer sind. Allerdings lassen sich für einfache Umladestrategien die Synergieeffekte durch den Güterumschlag abschätzen, was algorithmisch ausgenutzt werden kann (s. Abschnitt ??) Diese Ergebnisse flossen in einen lokalen Suchalgorithmus ein, der schon in seiner einfachsten Form für Probleme mittlerer Größe gute Ergebnisse bringt. Da

sich lokale Suchheuristiken leicht modifizieren lassen, eignet er sich auch hervorragend für die Anwendung in vielen praktischen Problemen. Deweiteren wurde im Rahmen einer Diplomarbeit ein *Column Generation*-Algorithmus für ein Anwendungsproblem aus der Automobilindustrie entwickelt.

Kontakt: `combopt@zpr.uni-koeln.de`

1.4 Steinerprobleme

Umladeprobleme spielen eine wachsende Rolle in der modernen Transportlogistik. Dabei wird bei einem Pickup-and-Delivery-Problem erlaubt, Güter kurzzeitig in einem sogenannten Konsolidierungszentrum zwischenzulagern und dann mit einem anderen Fahrzeug weiterzutransportieren. Dieser zusätzliche Freiheitsgrad erschwert jedoch die Modellierung erheblich.

Um ein tieferes Verständnis für die Eigenschaften solcher Probleme zu erhalten, haben wir zwei kombinatorische Optimierungsprobleme entwickelt. Diese sind zwar zu stark idealisiert, um direkt für die Anwendung nutzbar zu sein, wir denken jedoch, dass sie wichtige strukturelle Eigenschaften der Probleme abbilden. Zudem konnten wir die algorithmische Lösung eines der Probleme erfolgreich in einer Heuristik für die Anwendung einsetzen (s. Abschnitt ??).

Beim sogenannten *k-Star-Hub-Problem* gehen wir davon aus, dass jeder Auftrag einzeln direkt ausgeliefert wird oder alle Aufträge eines Kunden zu einem von k Konsolidierungszentren gebracht werden können. Dieses Problem lässt sich für nur ein Konsolidierungszentrum als bipartites Vertex Cover Problem modellieren, für $k = 2$ kann es mit Netzwerkflussmethoden ebenfalls effizient gelöst werden. Bei mehr als zwei Zentren ist es jedoch \mathcal{NP} -vollständig. Die zweite Art von Problemen haben wir *Steiner-Diagram-Problem* genannt. Das Problem kann auch als Umladeproblem interpretiert werden, bei dem jeder Auftrag auf seinem Weg beliebig oft umgeladen werden darf. Es ist verwandt mit dem generalisierten gerichteten Steinernetzwerk-Problem und dem minimalen äquivalenten Netzwerk-Problem. Bei dem behandelten Problem kommt zusätzlich die Forderung nach Kreisfreiheit hinzu, welche es uns zu zeigen erlaubt, dass sich das Problem in polynomieller Zeit lösen lässt, falls die Anzahl der Aufträge beschränkt, das zugrundeliegende Netzwerk transitiv abgeschlossen und die Dreiecksungleichung erfüllt ist. Die beiden letzten Bedingungen werden in einem Anwendungsproblem immer erfüllt sein.

Kontakt: `combopt@zpr.uni-koeln.de`

1.5 Ein Färbungsproblem aus der Automobilproduktion

Bei der Bestellung eines neuen Automobils haben Kunden die Wahl zwischen einer Vielzahl von Ausstattungsmerkmalen (Sonnendach, 3 oder 5 Türen, etc.) und Farben. Die europäische Automobilindustrie trägt dem Rechnung, indem verschiedene Modelle nicht – wie z. B. in den Vereinigten Staaten üblich – auf Vorrat, sondern der Nachfrage entsprechend produziert werden. Dies erfordert eine sorgfältige Planung der Produktion; insbesondere hat die Reihenfolge, in der eingehende Bestellungen bearbeitet werden, erheblichen Einfluss auf Qualität und Kosten. Wir beschäftigen uns mit einem Teilproblem des Produktionsprozesses, das in der Lackierstraße (dem *paint shop*) auftritt, in der täglich eine Sequenz von verschiedenen Karosserietypen in den nachgefragten Farben lackiert wird. Bei jedem Farbwechsel müssen die Farbdüsen der Sprühpistolen gereinigt werden. Dies erhöht einerseits die Produktionskosten, andererseits belasten die überschüssigen Farbreste das Abwasser. Eine Minimierung der Farbwechsel ist also wünschenswert.

Bislang werden dazu heuristische Verfahren eingesetzt. Dabei werden zumeist Farbsortierspeicher benutzt, mit denen die vorgegebene Sequenz von Karosserietypen kurzzeitig abgeändert und anschließend wieder hergestellt werden kann.

Wir konnten durch eine geeignete Abstraktion neue theoretische Aussagen über die einfachste Form des Problems gewinnen, bei der die Reihenfolge der Karosserietypen unverändert gelassen wird und nur durch die Änderung der Farbreihenfolge die Anzahl der Farbwechsel minimiert werden soll. Es zeigte sich, dass bereits dieses Problem sowohl für nur zwei verschiedene Karosserietypen und eine beliebige Anzahl von Farben als auch für nur zwei Farben und eine beliebige Anzahl von Karosserietypen \mathcal{NP} -vollständig ist. Für den Fall, dass sowohl die Anzahl der verschiedenen Karosserietypen als auch die Anzahl der Farben beschränkt ist, konnten wir ein dynamisches Programm angeben, das das Problem in polynomieller Zeit löst.

Kontakt: `combopt@zpr.uni-koeln.de`

1.6 Testmengen in der ganzzahligen Optimierung

Die Kenntnis von Testmengen ermöglicht es, ganzzahlige Optimierungsprobleme mittels eines primalen Optimie-

rungsalgorithmus zu lösen. Testmengen sind durch folgende Eigenschaft charakterisiert: Eine zulässige Lösung x eines ganzzahligen Optimierungsproblems ist entweder optimal, oder es gibt ein Element t der Testmenge, so daß die Differenz $x - t$ wieder zulässig ist und bezüglich der Zielfunktion einen besseren Wert liefert. Ist also eine zulässige Lösung gegeben, kann man iterativ durch Subtraktion von Testmengenelementen zur Optimallösung gelangen.

Conti und Traverso haben einen interessanten Zusammenhang zwischen ganzzahliger Optimierung und Gröbnerbasen in der kommutativen Algebra aufgezeigt. Dabei werden zulässige Lösungen einer bestimmten ganzen Familie von ganzzahligen Programmen in Beziehung zu einem Binomideal in einem Polynomring gesetzt. Bei geeigneter Wahl einer Monomordnung entspricht dann die reduzierte Gröbner-Basis dieses Ideals der minimalen Testmenge für die Problemfamilie. Theoretisch ergibt sich dadurch die Möglichkeit, mit Hilfe algebraischer Methoden ganzzahlige Optimierungsprobleme zu lösen: Erst wird mit dem Buchberger-Algorithmus die Testmenge ermittelt. Anschließend wird ausgehend von einer zulässigen Lösung eine Folge von Verbesserungsschritten „entlang“ Elementen der Testmenge ausgeführt, bis das Optimum erreicht ist.

Obwohl in dem betrachteten Fall die Gröbnerbasis nur zu Idealen mit recht spezieller Struktur berechnet werden muß, ist dieser Ansatz aufgrund der hohen Laufzeit des Buchberger Algorithmus und der potentiell gewaltigen Größe der Basis für Beispiele mit vielen Variablen sehr problematisch. Daher werden die algebraischen Methoden nicht als direkte Lösungsverfahren betrachtet. Statt dessen werden die Methoden der Computeralgebra im hier verfolgten Ansatz dazu eingesetzt, die minimalen Testmengen zu ganzzahligen Optimierungsproblemen zu berechnen, um anschließend aus den so gewonnenen Informationen Aussagen über die Struktur der Testmengen zu ermöglichen.

So ist es beispielsweise im Falle der Bestimmung einer Knotenüberdeckung minimalen Gewichts in einem Graphen gelungen, die Struktur der Testmenge vollständig zu beschreiben, und zwar unabhängig von der Tatsache, daß zu ihrer Berechnung aufwendige algebraische Methoden eingesetzt wurden.

Weiterhin wurde ein Vertex-Cover-Problem auf vollständigen Graphen betrachtet. Bei geeigneter Wahl der Monomordnung gilt in diesem Spezialfall der

Satz: Bis auf ein Element ist die reduzierte Gröbnerbasis genau die Menge aller Vektoren, die entweder negativ

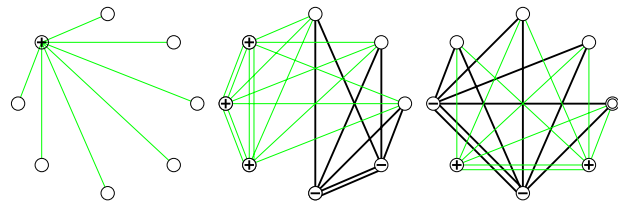


Abbildung 1.3: Typische Repräsentanten der minimalen Testmenge für den K_8 graphisch dargestellt. Das linke Element und Elemente vom mittleren Typ senken das Gewicht einer Knotenüberdeckung, während die Elemente vom rechten Typ lediglich für eine Umverteilung der Gewichte stehen.

bzgl. der Monomordnung sind und das Gewicht um 1 erniedrigen, oder das Gewicht unverändert lassen und positiv bzgl. der Ordnung sind.

Im Falle beliebiger Graphen kommt noch eine dritte Klasse von Elementen zur Testmenge hinzu, die die komplizierteren Adjazenz-Beziehungen zwischen Knoten mit veränderten Gewichten widerspiegelt.

Ziel ist es, bei schweren ganzzahligen Optimierungsproblemen eine – möglicherweise unvollständige – Beschreibung der Struktur der Testmenge anzugeben, deren Kenntnis dann in einer Verbesserungsheuristik genutzt werden kann.

Kontakt: combopt@zpr.uni-koeln.de

Kapitel 2

Forschungsschwerpunkt Verkehr

Neben interessanten wissenschaftlichen Fragestellung aufgrund der komplexen Dynamik von Verkehrssystemen ist die Beschäftigung mit diesen von hoher gesellschaftlicher Bedeutung. Gefragt sind wirkungsvolle Werkzeuge, die die Beurteilung von Verkehrszuständen und die vorausschauende Bewertung von Steuerungsmaßnahmen ermöglichen. Hierbei hat sich in den letzten Jahren die mikroskopische Simulation des Verkehrsgeschehens zu einem solchen Werkzeug entwickelt, das im Gegensatz zu konventionellen Modellen der Verkehrsplanung in der Lage ist, die wichtigen dynamischen Effekte im Verkehr abzubilden. Zu diesen Eigenschaften zählen das stark über Tag (rush hour), Woche und Jahr (Ferienverkehr) variierende Verkehrsaufkommen, sowie die Bildung von Staus und die Einzelfahrzeugdynamik.

Seit vielen Jahren bildet die Untersuchung und Entwicklung mikroskopischer Verkehrsmodelle einen bedeutenden Schwerpunkt in unserer Arbeitsgruppe. Das Arbeitsfeld reicht inzwischen von der Modellierung mobilitäts-erzeugender Verhaltensmuster und der Beschreibung der daraus resultierenden Verkehrsnachfrage, über die dynamische Routenumlegung bis zur eigentlichen Simulation der Bewegung der einzelnen Fahrzeuge und der Berechnung der sich ergebenden Schadstoffemissionen.

Unsere Arbeit findet in sehr enger Kooperation mit dem Deutschen Zentrum für Luft- und Raumfahrt statt, das seit Anfang dieses Jahres ein neues Institut mit dem Schwerpunkt Verkehrsforschung/-technik errichtet. Die dort ansässige Arbeitsgruppe ist 1996 aus unserer Gruppe entstanden und wächst seit Entstehung des neuen Forschungsschwerpunktes stetig. Neben dieser sehr fruchtbaren Zusammenarbeit ist unsere Arbeitsgruppe auch in vielen anwendungsorientierten Projekten (s.a. Abschnitt ??) mit unterschiedlichsten Projektpartnern aktiv. Dies ist ein ständiger Motor für die Weiterentwicklung unserer Modelle und die Erschließung neuer Bereiche aus dem Umfeld der Verkehrssimulation. Insbesondere hat sich die Ar-

beitsgruppe im vergangenen Jahr erstmals mit der Modellierung von Schienenverkehr beschäftigt.

2.1 Wissenschaftlicher Fortschritt

Innerhalb unserer anwendungsorientierten Forschungsprojekte ergeben sich ständig neue wissenschaftliche Fragestellungen, die im Rahmen von Diplom- und Doktorarbeiten untersucht werden. Neben daraus resultierenden zahlreichen Veröffentlichungen, wurden im Berichtzeitraum verschiedene Dissertationen und Diplomarbeiten abgeschlossen. Die Dissertationen beschäftigten sich inhaltlich mit den folgenden Themen:

Heuristische Verfahren zur Bestimmung von kürzesten Wegen in großen Straßengraphen

Das Problem, einen kürzesten Weg zwischen zwei Knoten in einem gewichteten Graphen zu finden, ist eines der klassischen Probleme in der Netzwerkoptimierung, das seit über vierzig Jahren Gegenstand ausgiebiger Forschungstätigkeit ist. In vielen praktischen Anwendungen taucht die Bestimmung von kürzesten Wegen auf, entweder als eigenständige Fragestellung (z.B. bei Transportproblemen, Projektmanagement und DNA-Sequenzierung) oder als Teilproblem eines komplexeren Problemzusammenhangs (z.B. bei der Approximation von Funktionen und dem Knapsack-Problem).

Im Rahmen einer Dissertation wurden verschiedene algorithmische Verfahren zur Bestimmung kürzester Wege

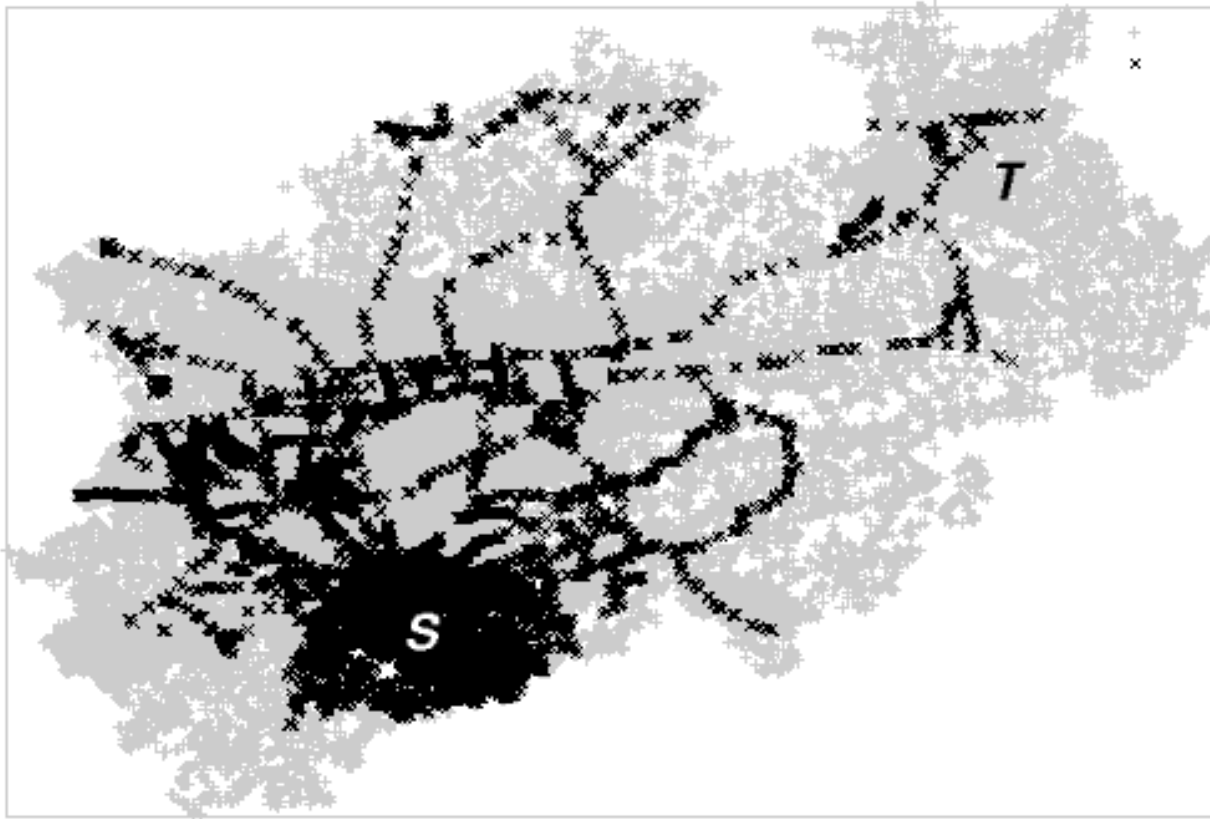


Abbildung 2.1: Bestimmung des kürzesten Weges zwischen den Knoten S und T. Es wird ein Rückwärtsdijkstra vom Zielknoten T aus auf dem Suchgraphen der Partitionsklasse, in der S liegt, durchgeführt. Im Gesamtgraphen sind alle Knoten schwarz eingefärbt, die während der Suche vom Algorithmus „besucht“ werden.

in sehr großen Straßengraphen untersucht. Die Beschäftigung mit diesen Fragestellungen wurde angeregt durch die zunehmende praktische Bedeutung der Generierung von geeigneten Routenempfehlungen für eine Vielzahl von Fahrern in Straßennetzen.

Ausgelöst wurde diese Entwicklung durch die Notwendigkeit einer effektiveren Verkehrslenkung aufgrund des immer weiter steigenden Verkehrsaufkommens und durch den Einsatz neuer Technologien im Bereich der Fahrzeugelektronik in den letzten Jahren. Hier sind insbesondere die zunehmende Verbreitung von individuellen Navigationssystemen und das damit verbundene Gebiet der Telematik zu nennen. Seit der Markteinführung solcher Navigationssysteme vor einigen Jahren verzeichnen die Hersteller jährlich sich verdoppelnde Umsatzzahlen mit einem erwarteten Absatz von über 600.000 Geräten in Deutschland für das Jahr 2000. Dabei tendiert die Entwicklung zu einer Routenführung hin, bei der zunehmend dynamische Verkehrsdaten berücksichtigt werden.

Auch bei einer effizienteren Planung und Steuerung von Verkehr mittels der am Institut entwickelten einzelfahrzeugbasierten Mikrosimulation müssen bei der iterativen Bestimmung des dabei angestrebten Verkehrsgleichgewichts idealerweise in jedem Iterationsschritt die Routen aller Fahrer neu berechnet werden. Bereits für ein relativ kleines Untersuchungsgebiet wie das Straßennetz der Stadt Wuppertal mit ungefähr 17.000 Kanten im Graphen bedeutet dies die Berechnung von 500.000 Routen pro Iterationsschritt, was selbst auf Computerworkstations der aktuellen Technik mehrere Stunden in Anspruch nimmt.

Den beschriebenen Anwendungen ist gemeinsam, dass eine benötigte Route nicht unbedingt der tatsächlich kürzeste Weg zwischen den beiden Endpunkten sein muss. Es reicht, wenn die Fahrtzeit einer vorgeschlagenen Routenempfehlung hinreichend nahe an derjenigen des optimalen Weges ist. Für die Zufriedenheit des Benutzers eines individuellen Navigationssystems ist es dabei insbesondere wichtig, dass seine persönlichen Erwartungen

erfüllt werden. Dagegen steht im Rahmen einer Verkehrssimulation eine ausreichende Beschreibung des realen Verkehrszustandes im Vordergrund. Dieser ist gegenüber vereinzelt nicht zu großen Abweichungen der Routen von optimalen Wegen relativ unempfindlich.

Für das darunter liegende Problem der Bestimmung von kürzesten Wegen in Graphen mit nicht-negativer Kostenfunktion ist der klassische Lösungsalgorithmus das bereits 1959 von Dijkstra vorgeschlagene Verfahren. Die schnellsten Implementierungen von Dijkstras Algorithmus benötigen für die Bestimmung eines kürzesten Weges im Straßengraphen von Nordrhein-Westfalen mit etwas über einer Million Kanten weniger als eine Sekunde auf einer Sun Enterprise E450 mit 4 mit 400 MHz getakteten UltraSPARC-II Prozessoren.

In praktischen Anwendungen kommen häufig heuristische Verfahren zur Routengenerierung in Straßengraphen zum Einsatz. Solche Methoden nutzen meist die spezielle Struktur von Straßengraphen mit gerichteten Kanten, Kantenlängen nah der euklidischen Distanz der beiden Endknoten sowie die Fastplanarität und eine hierarchischen Struktur aufgrund einer Typisierung der Kanten nach Wichtigkeit.

Im Rahmen der Dissertation wurde eine Heuristik entwickelt, die die Ähnlichkeit von kürzesten-Wege-Bäumen in Straßengraphen für nah beieinander liegende Startknoten ausnutzt. Die Heuristik gliedert sich in mehrere Phasen. Zuerst wird der Graph in k Klassen möglichst gleicher Größe partitioniert, und nachdem für jede Partitionsklasse eine Reihe von Basisknoten bestimmt wurde, wird in jeder der Klassen ein Suchgraph auf Basis dieser Knoten erzeugt. Diese Suchgraphen enthalten wesentlich weniger Kanten als der Gesamtgraph, aber alle Knoten des Graphen. Die Kantenmenge eines Suchgraphen besteht aus allen Kanten der Klasse und der Vereinigung aller Kanten von kürzesten-Wege-Bäumen der Basisknoten. Die Suchgraphen sind somit lokal dicht, aber global dünn besetzt. Die Anwendung eines Rückwärtsdijkstra bei der eigentlichen Routensuche führt schließlich zu einem sehr schnellen Algorithmus, s.a. Abbildung ??.

Auf den größten betrachteten Netzen ist die Baumheuristik um einen Faktor drei bis acht schneller als Dijkstras Algorithmus. In Bezug auf die Anzahl permanent markierter Knoten erzielt die Heuristik einen Gewinn um den Faktor 7 bis 20. Dabei werden Pfade gefunden, die im Mittel um weniger als 1% vom optimalen Weg abweichen, sofern die beiden Endknoten nicht zu nahe beieinander liegen. Es ist zu erwarten, dass über 90% der exakten kürzesten Wege von der Heuristik gefunden werden.

Darüber hinaus wurden etablierte Heuristiken untersucht und für das spezielle Problem auf Straßengraphen angepasst. Der Vergleich der Baumheuristik mit diesen Ansätzen zeigt, dass die neu entwickelte Heuristik aufgrund der gleichzeitigen Ausnutzung von geometrischen und hierarchischen Strukturen der Straßennetze und der Möglichkeit der dynamischen Anpassung der Suchgraphen zu sehr guten Ergebnissen führt.

Untersuchungen zur parallelen, objektorientierten Verkehrssimulation

Um Mehrprozessorsysteme mit verteiltem Speicher effizienter für die Simulation von Verkehr nutzen zu können, wurde innerhalb einer Dissertation untersucht, in wieweit sich objektorientierte Verkehrssimulationen und insbesondere die in unserer Arbeitsgruppe entwickelte Software PLANSIM-T parallelisieren lassen. Wesentlich hierzu ist die hinsichtlich der Lastenverteilung ausgewogene Zerlegung des zugrunde liegenden gewichteten Objektgraphen, der das Straßennetz mit seinen Eigenschaften repräsentiert. Die daraus resultierenden Teilgraphen werden dann parallel von unterschiedlichen Rechnern oder Prozessoren bearbeitet. Um die Kommunikation zwischen den einzelnen Prozessen niedrig zu halten, ist es nötig, die Anzahl der geschnittenen Kanten möglichst gering zu halten, da eine Absprache zwischen den Prozessoren genau dann erforderlich ist, wenn Fahrzeuge zwischen zwei Teilnetzen bewegt werden.

Aus diesem Grund ist der Algorithmus, der das Netz auf die Prozessoren aufteilt, besonders wichtig. Es wurden hierfür verschiedene Verfahren untersucht, die zur Zerlegung eines Objektgraphen verwendet werden können. Dabei wurden sowohl einfache Verfahren betrachtet, die keine Informationen über die Struktur des Graphen verwenden, als auch physikalisch motivierte sowie geometrische Verfahren, die Zerlegungen über die Berechnung von Trägheitsmomenten oder harmonischen Schwingungen finden. Zusätzlich wurden Reduktionsverfahren untersucht, die aus dem Graphen einen vereinfachten Hilfsgraphen konstruieren, der dann mit einem einfachen Verfahren zerlegt wird. Diese Zerlegung lässt sich mit lokalen

Optimierungsverfahren auf den ursprünglichen Graphen übertragen. Es hat sich gezeigt, dass als Reduktionsverfahren die von Karypis vorgeschlagen Multilevel k -way Partitionierung die für das vorliegende Problem geeignetesten Ergebnisse liefert. Dieses Verfahren lässt sich zudem sehr gut parallelisieren.

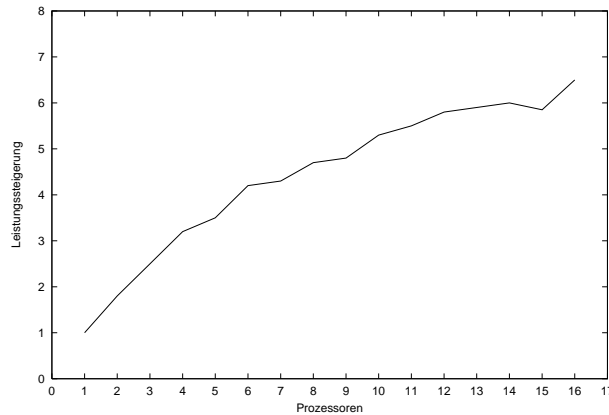


Abbildung 2.2: Erzielte Leistungssteigerung durch die Parallelisierung von PLANSIM-T mit zunehmender Zahl von Prozessoren.

Neben den theoretischen Überlegungen über die Zerlegung von Straßennetzen wurde eine universell einsetzbare Bibliothek entwickelt, welche es den einzelnen Prozessen ermöglicht, miteinander zu kommunizieren. Diese wurde speziell auf die Verkehrssimulation PLANSIM-T zugeschnitten und greift auf die frei erhältlichen Standardbibliotheken PVM (Parallel Virtual Machine) und MPI (Message Passing Interface) zurück, welche die Programmierung eines heterogenen Netzwerks von parallelen und seriellen Computersystemen als einen einzigen parallelen Rechner mit verteiltem Speicher ermöglicht.

Um die Leistungssteigerung zu messen, wurden verschiedene Netze auf unterschiedlichen Rechnersystemen simuliert und der Geschwindigkeitszuwachs bewertet. Abbildung ?? zeigt die Ergebnisse für ein großes Netz von NRW mit über 50.000 Kanten und 65.000 Knoten. Als Testrechner wurde eine SGI SC 900 XL mit 16 Prozessoren mit je 75 MHz und insgesamt 8 Gigabyte Hauptspeicher verwendet.

Es zeigt sich, dass die Leistung mit der Anzahl der Prozessoren steigt, dabei jedoch die Zuwachsrate sinkt. Diese Sättigungstendenz hängt mit der Kommunikation zwischen den Prozessoren zusammen, die mit zunehmender Anzahl immer aufwändiger und ineffizienter wird.

Abschließend lässt sich feststellen, dass durch die angewandten Verfahren zur Parallelisierung der Verkehrs-

flusssimulationen PLANSIM-T wesentliche Leistungssteigerungen erreicht werden können. Die erzielbare Steigerung ist dabei stark abhängig von der Struktur des zugrundeliegenden Straßennetzes. Insbesondere im Hinblick auf große Netze mit vielen Einzelfahrzeugen wurde die Einsatzfähigkeit der Software PLANSIM-T durch die Parallelisierung deutlich erweitert. Die Überlegungen der Arbeit lassen sich desweiteren auf andere Verkehrssimulationen übertragen, die objektorientiert konzipiert sind.

2.2 Anwendungsprojekte

3V – Verteilte Verkehrssimulation und Visualisierung

Das im August 1997 gestartete und inzwischen abgeschlossene Projekt Gigabit-Testbed West (GTB) wurde vom DFN eingeleitet und vom BMBF gefördert. Ziel dieses Projektes war die Erprobung und Untersuchung der Gigabit-Netztechnik, welche inzwischen vom DFN in Betrieb genommen wurde. Hierzu wurden verschiedene wissenschaftliche Anwendungen entwickelt, die sich die neue Netztechnik zu nutze machen sollten und sich ohne die im GTB vorhandene hohe Bandbreite nicht umsetzen ließen.

In dem von uns in Zusammenarbeit mit der GMD, dem DLR und der Firma echtzeit GmbH bearbeiteten Teilprojekt 3V, wurde die vom ZAIK entwickelte Simulationssoftware PLANSIM-T weiterentwickelt, um Möglichkeiten zu schaffen, die in Echtzeit errechneten Daten auf mehreren Rechnern, welche über das GTB miteinander verbunden sind, zu visualisieren. Dies war aufgrund des hohen Datenaufkommens der Mikrosimulation von etwa 400 Mbits/s bisher nicht möglich.

Durch die Trennung von Simulation und Visualisierung sowie der Option, verschiedene Visualisierungen an unterschiedlichen Arbeitsplätzen betreiben zu können, ergeben sich vielfältige neue Nutzungsmöglichkeiten. So sind von einer einfachen zweidimensionalen, über verschiedene wissenschaftliche Darstellung bis hin zur komplexen VR-Visualisierung verschiedene Darstellungsformen denkbar, von denen zwei innerhalb des Projektes realisiert wurden. Durch die Unabhängigkeit der Visualisierung von der Simulation lässt sich diese einfach auf die Bedürfnisse verschiedener Benutzergruppen zuschneiden. Heterogene Gruppen können räumlich voneinander getrennt durch die

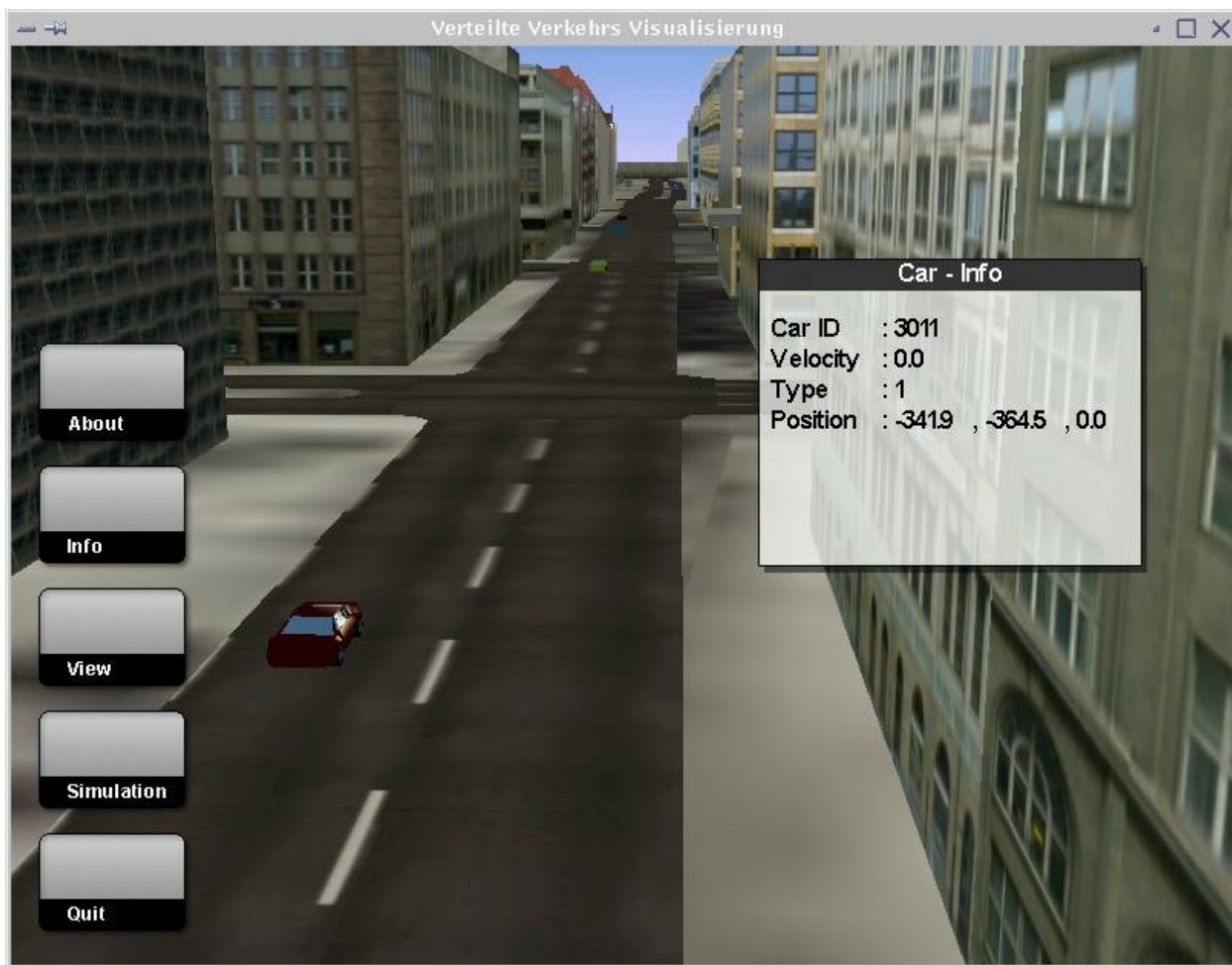


Abbildung 2.3: 3D-Visualisierung (echtzeit GmbH) der von PLANSIM-T erzeugten Daten.

Nutzung der Datenübertragung im GTB zusammenarbeiten. Denkbar wäre die gemeinsame Erarbeitung und Bewertung von verkehrlichen Maßnahmen durch verschiedene Entscheidungsträger aus Politik, Stadtplanung und Wissenschaft, indem den Gruppen simultan, die aus der Simulation einmal erzeugten Informationen, benutzergerecht präsentiert werden.

Eine weitere wichtige Funktion, um die PLANSIM-T erweitert wurde, ist die Möglichkeit der Interaktion zwischen dem Anwender der Visualisierung und der Simulation. Mit Hilfe eines Rückkanals können verschiedene Befehle an die Simulation übermittelt werden, die dann sofort im aktuellen Simulationsprozess ausgeführt werden. Die Anwendungsbereiche dieser Option reichen von einer einfachen Aufforderung an die Simulation, Daten über das Netz zu verschicken, bis hin zu einem direkten Eingreifen in die Simulation und die Ausführung vorgefertigter Szenarien.

Die Übertragung der Daten über das GTB wird mit Hilfe der multicastfähigen Netzprotokolle IP-Multicast und ATM vorgenommen. Dadurch muss der Datenstrom nur ein einziges mal durch den Simulator erzeugt werden, und die Netzbelastung lässt sich so auch für mehrere angeschlossene Visualisierungen konstant halten. Während IP-Multicast sehr weit verbreitet und auf nahezu jedem modernen Rechner verfügbar ist, bietet ATM den Vorteil, eine gewisse Bandbreite für die Daten reservieren zu können. Es kann bei der Verwendung von ATM also gewährleistet werden, dass die gewünschten Daten auch rechtzeitig bei der Visualisierung ankommen. Für den Rückkanal wird zwischen jeder Visualisierung und der Simulation eine eigene TCP/IP-Verbindung hergestellt.

Zur Demonstration der erzielten Ergebnisse wurde ein Teil des Berliner Stadtkerns in ein Beispielszenario übernommen und simuliert. Die daraus gewonnenen Daten ließen sich auf verschiedenen über das Gigabit-Testbed West

angeschlossenen Rechnern darstellen. Dies geschah mit Hilfe der im Rahmen des Projektes entwickelten wissenschaftlichen 2D-Visualisierung und der von der Firma echtzeit GmbH entwickelten 3D-Visualisierung (siehe Abbildung ??).

Die Visualisierungen konnten sich zu jeder Zeit während des Simulationslaufes mit Hilfe des Rückkanals an die Simulation an- und wieder abmelden. Um die weiteren Fähigkeiten des Rückkanals aufzuzeigen, wurde ein Szenario erstellt, in dem die Sperrung einer Straße vom Anwender innerhalb der Visualisierung durchgeführt werden kann. Die Verkehrsströme zwischen den Quellen und Senken werden hierfür in Form einer Start/Zielmatrix vorgegeben und die Fahrzeuge suchen sich in Abhängigkeit vom momentanen Verkehrszustand den kürzesten Weg. Sperrt man nun während des Simulationslaufes die gewählte Straße, so kann man die Auswirkungen dieser Maßnahme direkt mit Hilfe der angeschlossenen Visualisierungen analysieren.

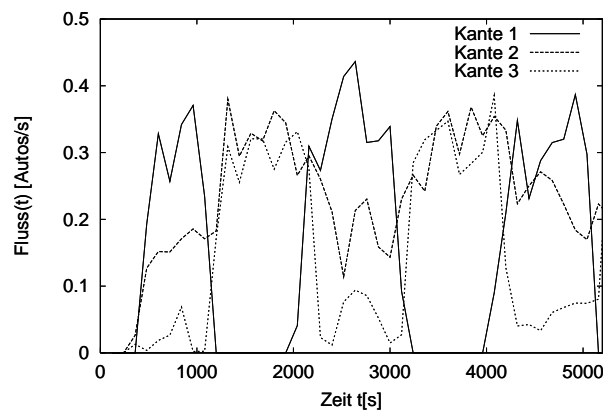


Abbildung 2.4: Gemessener Fluss auf drei Kanten des Systems für das Beispielszenario „Straßensperrung“. Kante 1 wird zeitweise gesperrt, die anderen beiden stellen Ausweichstrecken dar.

In einem Beispielszenario wurde eine Straße alle 1000 Zeitschritte für 500 Zeitschritte gesperrt. Abbildung ?? zeigt den zeitlichen Verlauf des Flusses auf drei Kanten des Systems. Kante 1 ist diejenige, die gesperrt wird. Die Kanten 2 und 3 liegen auf Alternativrouten der betrachteten Start/Ziel-Beziehung. Deutlich erkennt man den schnellen Abfall des Flusses während der Sperrzeiten, sowie die Verteilung des Verkehrs auf die Ausweichstrecken.

Es lassen sich leicht weitere Beispiele entwickeln, anhand deren man die Vorteile der Trennung zwischen Simulation und Visualisierung sowie des Rückkanals zeigen

kann. Nicht zuletzt die Möglichkeit den Verlauf einer Simulation zwischendurch optisch kontrollieren zu können, ohne eine dauerhafte Visualisierung (welche die Simulationsgeschwindigkeit nur vermindern würde) zu betreiben, machen die im 3V-Projekt durchgeführten Arbeiten PLANSIM-T zu einem mächtigen Instrument zur Verkehrsflusssimulation.

Weitere Informationen über das Projekt 3V finden sich auch auf unseren Internetseiten.

Sonderforschungsbereich „Umweltprobleme eines industriellen Ballungsraumes“

Das ZAIK arbeitet im Sonderforschungsbereich 419 (SFB) mit dem Institut für Geophysik und Meteorologie der Universität zu Köln seit Anfang 1999 zusammen. Gegenstand der Untersuchungen in diesem Projekt sind die Entstehung, Ausbreitung und Transformation von Luftschadstoffen. Zusätzlich sollen hierbei auch klimatologische Fragestellungen berücksichtigt werden. Gesamtziel des SFB ist eine umfassende Untersuchung der Umweltproblematik – insbesondere der Verflechtungen zwischen den verschiedenen Problemkreisen Luft, Wasser, Boden. Auf der Basis von naturwissenschaftlichen und medizinischen Analysen, Simulationen und Prognosen unter Berücksichtigung von sozialen, ökonomischen und rechtlichen Aspekten soll diese zu Lösungsansätzen führen, Lebensqualität zu sichern, zu erhalten und zu verbessern. Mit den Untersuchungen zur Luftqualität trägt unser Projekt einen Teil zu dieser Gesamtsicht bei.

Die beiden Arbeitsgruppen haben bereits erfolgreich im 1998 ausgelaufenen Forschungsverbund „Verkehrssimulation und Umweltwirkungen“ kooperiert. Die im Bereich der Verkehrssimulation zur Untersuchung der Fragestellung verwendeten Modelle und Softwaretools basieren auf Forschungsergebnissen, die im Rahmen des Forschungsverbundes erzielt wurden. Dabei handelt es sich um Modelle zur dynamischen Routenumlegung und der mikroskopischen Simulation von Straßenverkehr. Diese werden im Rahmen des SFB-Projektes methodisch erweitert und ausgebaut.

Grundlage der in Abbildung ?? dargestellten Modellkette zur Simulation von Luftschadstoffen sind neben Emissionskatastern für Haus- und Industriebrand zeitlich und

Numerische Simulation

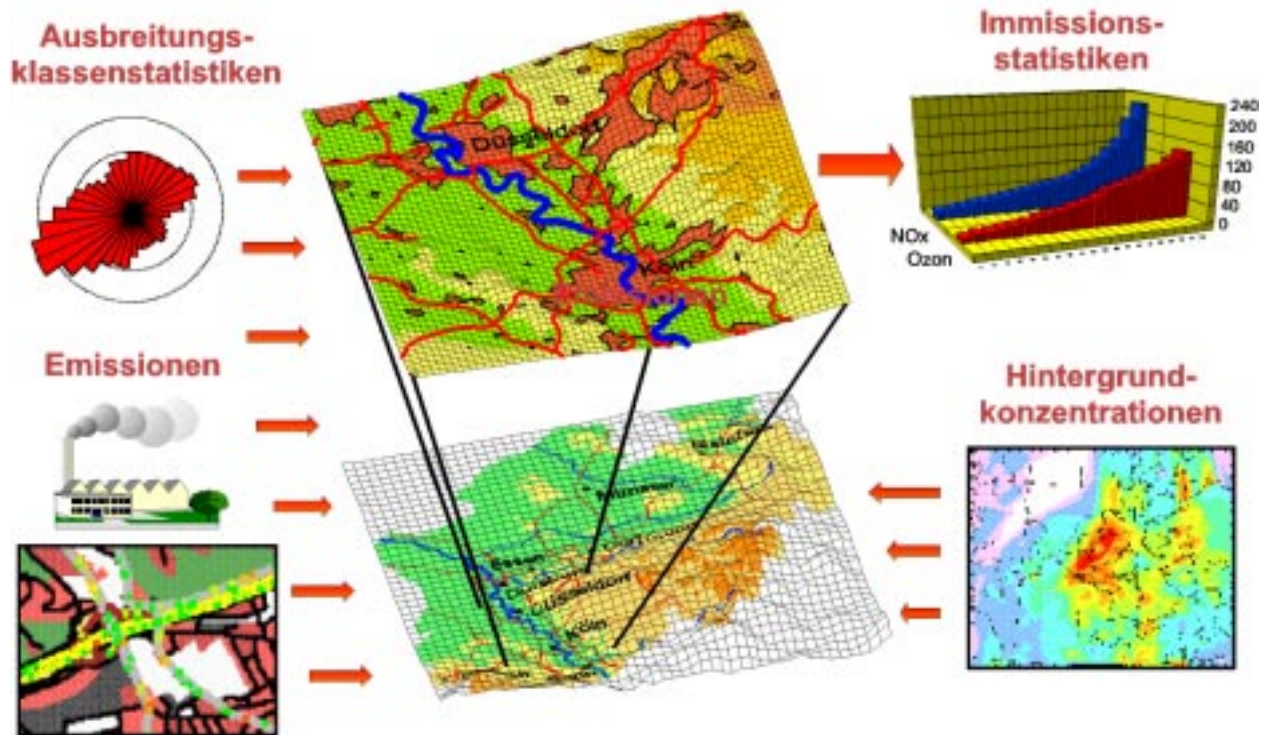


Abbildung 2.5: Modellkette des SFB 419 - Teilprojekt B3. Auf der Basis von Emissionskatastern von Industrie und Verkehr wird mit Hilfe hochauflösender meteorologischer Modelle die Ausbreitung und chemische Transformation der Schadstoffe (Immission) berechnet.

räumlich hochaufgelöste Kataster der Schadstoffemissionen des Straßenverkehrs, die mit Hilfe der in unserer Arbeitsgruppe entwickelten Modelle berechnet werden. Neben der naheliegenden Feststellung, dass in der Regel eine Erhöhung der Auflösung die Qualität der Ergebnisse verbessert, ist eine hohe Auflösung der Dynamik des des Straßenverkehrs für die Berechnung der Emissionen von prinzipieller Bedeutung. Die in einem Bereich des Straßennetzes emittierten Schadstoffe hängen nicht allein von der mittleren Verkehrsbelastung, sondern vielmehr von deren zeitlichen Entwicklung bis hin zur Einzelfahrzeugdynamik ab (s.a. Abbildung ??). Eine hohe räumliche Auflösung der berechneten Schadstoffkataster ist darüber hinaus notwendig, um geeignete Datenbasen für weitere Arbeitsgruppen des SFB zur Verfügung stellen zu können.

Die bisher im Rahmen des SFB 419 durchgeführten Arbeiten unserer Arbeitsgruppe konzentrierten sich stark auf die Generierung der Verkehrsnachfrage und die Koppelung der drei Komponenten der Modellkette, sowie die Erstellung des zur Simulation notwendigen Verkehrsnet-

zes. Durch die Bereitstellung eines umfangreichen Datensatzes der GIS-Consult GmbH (Haltern) konnte das gesamte Straßennetz von NRW für die Simulation des Verkehrsgeschehens abgebildet werden. Die Detaillierung der Netzdaten reicht hierbei bis auf Kreisstraßenniveau. Aufgrund der momentan zur Verfügung stehenden Daten über die Verkehrsnachfrage und der großen Skala, auf der die Emissionen des Verkehrs berechnet werden sollen, stellt dies zunächst auch die anvisierte Auflösung der Verkehrsabläufe dar. Aus der Zusammenarbeit mit der GIS-Consult GmbH sind zwei Diplomarbeiten im Bereich der Geowissenschaften entstanden, die die Darstellung dynamischer Verkehrsdaten, wie sie z.B. aus Simulationsläufen stammen können, in einer auf Smallworld basierenden Applikation (NWSIB – Straßeninformationsdatenbank Nordrhein-Westfalen) ermöglichen.

Grundlage jeder Verkehrssimulation ist die Kenntnis über die Verkehrsnachfrage. Flächendeckende Informationen sind in der Regel nicht verfügbar und liegen meist – wenn überhaupt – in einer stark aggregierten Form vor. Um die

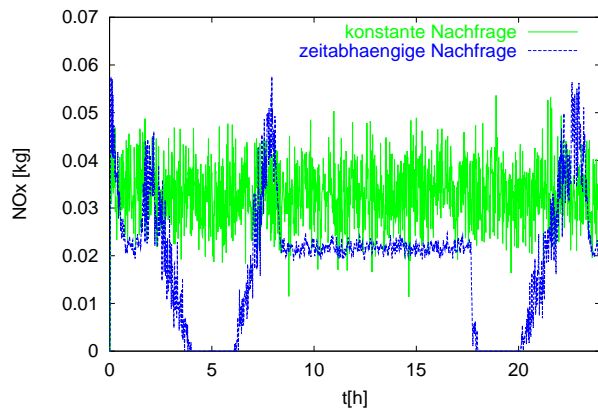


Abbildung 2.6: Vergleich der NO_x Emission auf einer Straßenkante für konstante und zeitabhängige (zwei rush-hour Peaks) Nachfrage. Nicht nur die zeitliche Verteilung des Schadstoffs unterscheidet sich, sondern auch die Gesamtemission ist im Falle konstanter Nachfrage ca. doppelt so hoch wie bei der zeitabhängigen (trotz gleicher Anzahl von Emittenden).

Modellkette innerhalb der Verkehrssimulation, d.h. dynamische Routenumlegung und Simulation der Einzelfahrzeugdynamik, weiterentwickeln zu können und möglichst schnell ein zeitlich hochaufgelöstes Emissionskataster für die Modelle der Meteorologen zur Verfügung zu stellen, wurde in einem ersten Schritt allein aufgrund der Bevölkerungszahlen der Gemeinden eine Verkehrsnachfrage um Köln abgeschätzt. In dem gewählten Ansatz fließen neben den Einwohnerzahlen die durch die Infrastruktur vorgegebenen Erreichbarkeiten zwischen den Gemeinden mit in die Verteilung der Fahrten ein. Basierend auf dieser Nachfrage wurde die in Abbildung ?? dargestellte NO_x Immission für Köln und Umgebung berechnet.

Nachdem eine Großzahl der technischen Probleme zur Umsetzung der Modellkette, wie die Kopplung der einzelnen Komponenten, inzwischen gelöst wurden, sind die Arbeiten momentan auf die Modellierung der Verkehrsnachfrage als Grundlage einer wirklichkeitsnahen Berechnung der Emissionen ausgerichtet. Als Datenbasis dienen hierbei die Verkehrsströme zwischen Landkreisen. In den vorliegenden Daten sind diese differenziert nach sechs Fahrtzwecken, wie Arbeit, Ausbildung oder Freizeit, enthalten. Mittels eines Gravitationsansatzes werden die Fahrten über ein iteratives proportionales Fitting auf Gemeinden verteilt, wobei deren Einwohnerzahl und die Verteilung der Reisezeiten zwischen den Gemeinden mitberücksichtigt werden. Die Dynamisierung der Fahrten geschieht anhand von Verteilungen der Abfahrtszeiten, die aus Vorar-

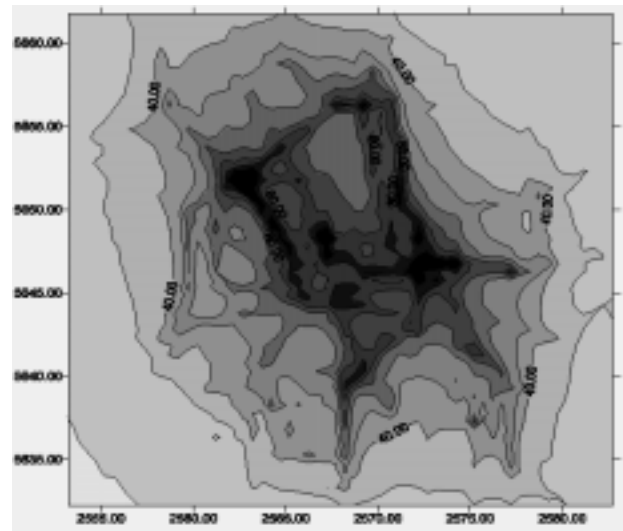


Abbildung 2.7: Mittlere NO_x Immission für die Umgebung von Köln. Simulation des Instituts für Geophysik und Meteorologie der Universität zu Köln aus Emissionsdaten basierend auf der im Text beschriebenen vereinfachten Nachfrageberechnung.

beiten eines anderen Projekts unserer Arbeitsgruppe gewonnen werden konnten. Das Verfahren wird zur Zeit für Fahrten des Berufsverkehrs getestet und soll anhand von Werten des durchschnittlichen Tagesverkehrs aus Studien des Umweltamtes der Stadt Köln validiert werden.

SimVV – Mobilität verstehen und lenken

In Zusammenarbeit mit dem ISB Aachen, dem DLR und dem Wuppertal Institut startete im Januar 2000 das vom Ministerium für Schule, Wissenschaft und Forschung des Landes Nordrhein-Westfalen (MSWF) geförderte Projekt SimVV, welches bis zum Ende des Jahres 2001 läuft. In diesem Projekt soll eine Modellkette entwickelt werden, die es ermöglicht, die Einwohner einer Stadt, sowie den Entwurf und die Durchführung der Tagespläne einzelner Personen zu simulieren. Die Notwendigkeit eines solchen Ansatzes liegt in der Tatsache begründet, dass bei der Berechnung der Verkehrsnachfrage bisher weitgehend auf die Modellierung der mobilitätsbestimmenden Faktoren und deren Wechselwirkungen verzichtet wurde. Aus diesem Grund lassen sich viele Fragen nach dem Verkehrsaufkommen und Möglichkeiten dieses zu beeinflussen nicht oder nur unzureichend beantworten. Um

die Ursachen von Mobilität zu erforschen und quantitative Schlussfolgerungen und Prognosen zu ermöglichen, müssten deshalb idealerweise die Entscheidungsprozesse von Einzelpersonen modelliert werden. In dem Projekt SimVV soll ein erster Schritt in diese Richtung unternommen werden.

Hierzu werden zunächst schon vorhandene Modelle zur Verkehrsnachfrageberechnung, der Aktivitätenmodellierung und Modelle der mikroskopischen Verkehrsfluss-Simulation verfeinert und in ein Gesamtwerkzeug integriert. Mit diesem soll die Machbarkeit, Mobilitätsursachen durch ein solches Vorgehen modellieren zu können, an einem konkreten Beispiel geprüft werden.

Die konzeptionellen Arbeiten an der Methodik zum Ablauf der Simulation sind inzwischen weitestgehend abgeschlossen, so dass jetzt mit der software-technischen Umsetzung begonnen wird. Im Folgenden wird das bisher erarbeitete Schema des Simulationsablaufes vorgestellt, s.a. Abbildung ??.

In einem ersten Schritt wird aus soziodemographischen Daten eine synthetische Bevölkerung generiert. Jede Person dieser Bevölkerung bekommt eine Auswahl von Aktivitäten mit Prioritäten, Dauer, möglichen Zeiten und Orten, sowie weitere Informationen zugewiesen. Im Rahmen des Modellierungsansatzes wählt der Akteur seine Handlungen aus den gegebenen Möglichkeiten aus und stellt aus diesen Aktivitäten einen Tagesplan zusammen. Hierbei werden persönliche und haushaltsspezifische Rahmenbedingungen wie verfügbares Geld, nutzbare Verkehrsmittel und Vorlieben in der Verkehrsmittelwahl berücksichtigt.

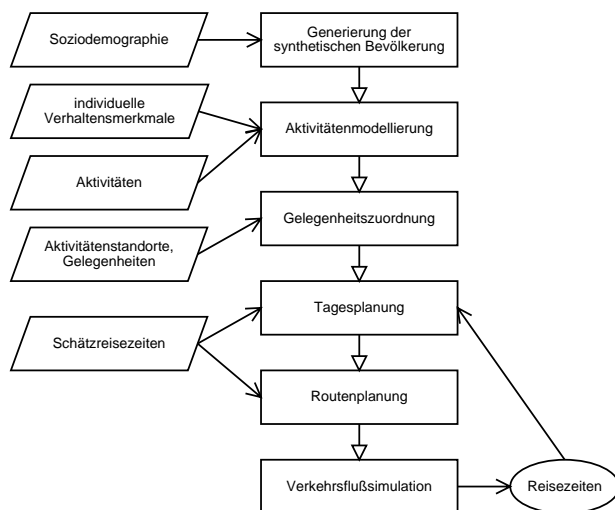


Abbildung 2.8: Gesamtkonzept SimVV

Ganz wesentlich für die Planung sind insbesondere die von den Akteuren für die Durchführung der Aktivitäten erwarteten Reisezeiten zwischen den einzelnen Gelegenheitsstandorten. Nachdem jeder Akteur seinen individuellen Tagesablauf geplant hat, werden die Pläne aller Akteure simultan im Verkehrsnetz durchgeführt. Hierbei wird das in der Arbeitsgruppe entwickelte Modell zur dynamischen Verkehrsumlegung eingesetzt. Die in der Simulation ermittelten Reisezeiten stellen eine wesentliche Information innerhalb der Modellkette dar, die im Rahmen der Modellierung allen Akteuren zugänglich ist. Diese Erfahrungen über die Reisezeiten im Verkehrsnetz nutzen diese dazu, die Tagespläne anhand ihrer Erfüllungbarkeit zu bewerten und neue Pläne zu generieren. Dieser Iterationsprozess wird solange fortgesetzt, bis das System in einen Gleichgewichtszustand läuft. Hinter diesem Vorgehen steht die Annahme, dass die beobachteten Mobilitätsmuster das Ergebnis solcher individuellen Lernprozesse sind.

Neuartig an dem hier skizzierten Modellablauf ist die Verknüpfung der aktivitätsbasierten, einzelpersonenbezogenen Verkehrsnachfrageberechnung mit der dynamischen Simulation der Bewegungsmuster von Menschen. So entsteht ein Instrumentarium, das zu quantitativen Aussagen über die Wirkungen und Wirksamkeit von mobilitätslenkenden Maßnahmen befähigt.

Ausgehend von den Arbeiten und Erfahrungen aus diesem Projekt sind eine Reihe von weitreichenden Erweiterungen unserer Modelle geplant. Insbesondere soll in der Zukunft intermodales Verhalten, d.h. die Benutzung mehrerer Verkehrsmittel, in den Modellen abgebildet werden. Die zum Individualverkehr alternativen Verkehrsmittel wie Bus und Bahn, Taxen, Mitfahrgelegenheiten spielen eine wichtige Rolle bei der Untersuchung des Mobilitätsverhaltens von Personen. Desweiteren gilt es einen Weg zu finden, wie Personenwirtschafts-, Durchgangs- und Fernverkehre mit in die Modellierung einfließen können, da diese nicht als Akteure in der Tagesplangenerierung auftauchen.

2.3 Simulation von Bahnanlagen

Ausbaumaßnahmen innerhalb von bestehenden Bahnanlagen stellen ein komplexes Planungsproblem dar. So sollte in einem gemeinsamen Projekt mit dem Ingenieurbüro K+V GmbH (Greven) die Frage untersucht werden, inwieweit die Errichtung einer Instandsetzungs- und Wartungsanlage den bisherigen Zugbetrieb innerhalb eines Bahnhofes beeinträchtigt, bzw. inwiefern die Betriebsabläufe

trotz der damit verbundenen Kapazitätsverringerung der verfügbaren Gleise durch einfache Modifikationen aufrechterhalten werden könnten. Wichtig ist diese Fragestellung zum einen, da der betrachtete Bahnhof aufgrund hoher Verkehrsnachfrage insbesondere zu Ferienzeiten stark ausgelastet ist. Zum anderen müssen die im laufenden Betrieb ständig anfallenden Instandhaltungsmaßnahmen außerhalb der Anlage durchgeführt werden, was zu zusätzlichen Fahrten und damit Kosten führt. Die Errichtung einer solchen Anlage innerhalb des Bahnhofsbereiches kann daher wesentliche Vorteile mit sich bringen.

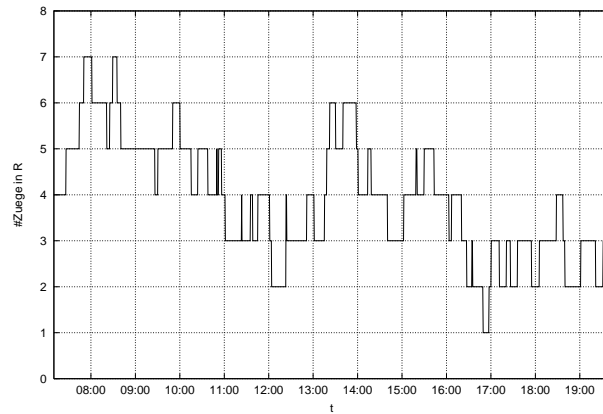


Abbildung 2.9: Zeitabhängige Auslastung eines Moduls mit maximaler Kapazität 8.

Zur Beantwortung dieser Frage wurde ein Simulator erstellt. In dem hierfür entwickelten Modell wird die Gleisanlage in Abschnitte, Module, zerlegt, die im wesentlichen durch ihre Kapazitäten charakterisiert werden. Die Züge werden anhand ihrer vorgegebenen Fahrpläne durch diese Module bewegt. Dabei agieren die Module über einen Satz von Regeln miteinander, die steuern, ob Züge das jeweilige Modul verlassen dürfen oder nicht. Sie garantieren damit zum einen, dass die Kapazitäten der Module nicht überschritten werden. Zum anderen werden so Kollisionen zwischen Zügen und artifizielle Blockadesituationen ausgeschlossen. Desweiteren ist es innerhalb des Modells möglich, wichtige Betriebsabläufe wie Blockabstände, Freimeldezeiten sowie das Koppeln und Entkoppeln von Zügen abzubilden.

Neben der Modularisierung der Anlage mit ihrem Regelwerk bildet der aktuelle Fahrplan die Basis der Simulation. Er stellt eine fahrbare, da bereits praktizierte Lösung in der Ausgangssituation dar und soll durch die Kapazitätsveränderung möglichst wenig beeinträchtigt werden, da der Fahrplan in überregionalem Zusammenhang gesehen werden muß. Der Fahrplan ist im Modell allerdings nicht

starr vorgegeben, sondern es werden lediglich Ankunfts- und Abfahrtszeit an den Bahnsteigen fest kodiert. Für die weitere Bewegung durch die Anlage sind Mindestverweildauern und frühestmögliche Abfahrtszeiten festlegbar, die notwendige Fahrzeiten und Servicezeiten garantieren. Unter Berücksichtigung der Vorgaben werden die Züge innerhalb des Modells durch die Anlage bewegt und notwendige Verschiebungen in den Plänen vorgenommen.

Die Ausgabe des Modells besteht aus den realisierten Umläufen der Züge durch die modularisierte Anlage sowie der Belegung der einzelnen Bereiche im Verlauf der Simulation (Abbildung ??). Die realisierten Umläufe werden tabellarisch den ursprünglichen Fahrplänen gegenübergestellt und aufgetretene Abweichungen kenntlich gemacht. So ist es zum einen für jeden Zeitpunkt möglich, ein aktuelles Bild der gesamten Anlage zu erhalten. Zum anderen ist unmittelbar ersichtlich, inwieweit die vorgegebenen Fahrpläne eingehalten werden können und an welchen Stellen Veränderungen im Betriebsablauf notwendig sind. Insofern Kapazitätsrestriktionen keine umfassende Umgestaltung der Betriebsabläufe notwendig machen, stellt der realisierte Fahrplan eine mögliche Lösung dar, die nahe am ursprünglichen Umlaufplan der Züge liegt. Durch die flexible Gestaltung der Eingabe können ohne großen Aufwand verschiedene Szenarien durchgespielt werden, um geplante Maßnahmen zu bewerten.

Mit Hilfe des Modells und des entwickelten Softwaretools konnte in dem gemeinsamen Projekt mit der K+V GmbH gezeigt werden, dass bei den vorliegenden betrieblichen Abläufen innerhalb des untersuchten Bahnhofs, die Errichtung der vorgesehenen Wartungsanlage nicht realisierbar ist. Die hierfür entwickelte Methode ist gut auf andere Fragestellungen dieser Art übertragbar und kann als Grundlage für weitere Optimierungsansätze dienen.

Kontakt: traffic@zpr.uni-koeln.de

Kapitel 3

Bioinformatik

Die moderne Molekularbiologie, ihre Umsetzung in der Forschung und Entwicklung, sowie die Bioinformatik als Motor des Fortschritts finden inzwischen selbst im Feuilleton von Tageszeitungen regelmäßig ihren Niederschlag. Die Biotechnologie ist inzwischen einer der wichtigen wirtschaftlichen Wachstumsfaktoren, sowie Inhalt zahlreicher Studiengänge an deutschen Universitäten.

Bemerkenswert bei dieser Entwicklung ist die Tatsache, dass z.B. die Erfolge von Celera Genomics beim viel publizierten Wettstreit um die Erstsequenzierung des menschlichen Genoms letztendlich Erfolge bioinformatischer Natur sind. Der wesentliche Schritt war, neben einem Fortschritt bei sogenannten Sequenzierungsrobotern, die für das Projekt einen finanziell überschaubaren Rahmen gewährleisteten, die Entwicklung eines Assembleralgorithmus. Dieser „Assembler“ muss die mit dem für die Sequenzierung benutzten Shotgun-Verfahren erzeugten kleinen Teilstücke der genomischen Sequenzen trotz hoher Fehlerraten in der richtigen Reihenfolge zur Gesamtsequenz zusammensetzen.

Anders als oft in den Medien dargestellt ist mit der Erstsequenzierung des menschlichen Genoms noch lange nicht das Ende der biotechnologischen Entwicklung erreicht. Die Sequenz alleine bedeutet noch keinen entscheidenden Erkenntnisgewinn. Vielmehr ist das Finden von Genen, die Bestimmung der Struktur und Funktion von Proteinen, das Erkennen und Verstehen von metabolischen Netzwerken nötig, also die systematische Aufklärung molekularer Mechanismen komplexer Lebensvorgänge.

Die Bioinformatik, als ein neues, eigenständiges, interdisziplinäres Arbeitsfeld, in dem Methoden der Mathematik, Informatik und Statistik benutzt werden, wird dabei auch weiterhin ein zentraler Baustein sein.

Am ZAIK wird seit 1996 an Fragestellungen der Bioinformatik geforscht. Inzwischen bestehen Kooperationen mit dem Los Alamos National Laboratory, den Instituten für Biochemie (AG Prof. Schomburg) und Genetik (AG Prof. Tautz) an der Universität zu Köln, dem

Max-Planck-Institut für Züchtungsforschung, sowie dem Bioinformatik-Startup Science Factory, ebenfalls in Köln.

In Lehre und Ausbildung wird durch Seminare und Vorlesungen zur Bioinformatik, durch enge, regelmäßige Kontakte auf Mitarbeiterebene und die von Mitarbeitern der Arbeitsgruppe ins Leben gerufene Bioinformatik-Interessen-Gruppe, einem Zusammenschluss von Studenten, Firmenangestellten und Wissenschaftlern, ein multidisziplinäres Umfeld für weitere Erfolge in der Bioinformatik aktiv gestaltet.

3.1 Ein neues Verfahren zum Lernen von Hidden-Markov-Modellen

Hidden-Markov-Modelle (HMM) sind eine Klasse von stochastischen Modellen, die schon für eine Vielzahl von Anwendungsgebieten, von der Spracherkennung bis zur Simulation biochemischer Prozesse in Zellmembranen, erfolgreich eingesetzt wurden. Am ZAIK werden z.B. HMM auch für die Analyse und Simulation ökonomischer Zeitreihen, genauer von Bausparverträgen, benutzt.

HMM bestehen aus einer Markov-Kette, also einer „gedächtnislosen“ Folge von Zufallsvariablen, den sogenannten Zuständen, die aber nicht direkt zu beobachten sind. Weiterhin gibt es für jeden dieser Zustände eine diskrete oder kontinuierliche Verteilung, anhand derer dann die beobachtbaren Ausgaben erzeugt werden. Ein HMM impliziert eine Wahrscheinlichkeitsverteilung auf der Menge aller möglichen Sequenzen von Ausgaben.

Für DNA-Sequenzen, also Abfolgen der Buchstaben A, C, G und T, kann in einem einfachen Beispiel ein Zustand des HMM jeweils einer funktional ausgezeichneten Region des DNA-Moleküls entsprechen (coding vs. non-coding). Diese unsichtbaren Zustände zeichnen sich

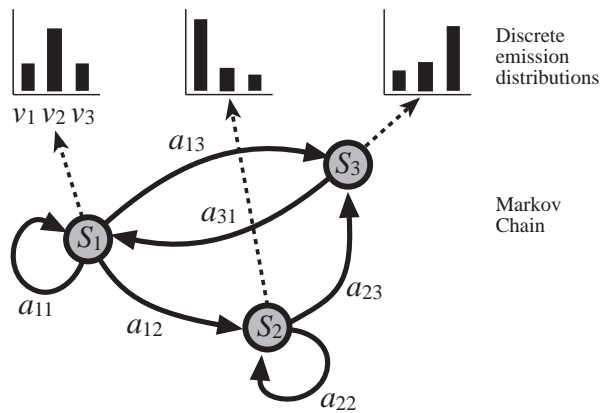


Abbildung 3.1: Ein HMM mit drei Zuständen als gerichteter Graph mit diskreten Ausgabeverteilungen. Die Kantenengewichte sind als Übergangswahrscheinlichkeiten zwischen Zuständen zu interpretieren.

z.B. durch unterschiedliche Sequenzkompositionen, also Unterschiede in der beobachteten relativen Häufigkeit der einzelnen Buchstaben A, C, G und T zwischen den verschiedenen Regionen aus. Ein mit entsprechendem Datenmaterial trainiertes HMM, kann dann z.B. für die Erkennung von kodierenden Regionen benutzt werden.

Weitere Anwendungsfelder von HMM in der Bioinformatik sind das Finden von Genen, Strukturvorhersagen, Verfahren zur Motivsuche, Finden entfernt homologer Proteinsequenzen, Multiple Alignments, Klassifizierung von DNA/Proteinsequenzen (Transkriptionseinheiten) und Modellierung von Proteinfamilien.

In der Literatur zu HMM, insbesondere in dem bekannten Tutorial-Artikel von Rabiner, werden drei Probleme in Bezug auf HMM erwähnt:

1. Wie kann man die Wahrscheinlichkeit ausrechnen, dass ein bestimmtes Modell eine gegebene Ausgabe-Sequenz ausgerechnet hat?
2. Gegeben ein Modell und eine Ausgabe-Sequenz, wie kann man die „dazugehörige“ Sequenz an „versteckten“ Zuständen rekonstruieren?
3. Wie kann man die Parameter eines HMM so trainieren, dass die Wahrscheinlichkeit, eine bestimmte Ausgabe-Sequenz zu erzeugen, maximiert wird?

Da die von Rabiner betrachteten Anwendungen in der Spracherkennung auf natürliche Art und Weise ein geeignetes HMM nahelegten, ist es verständlich, dass eine viel grundlegendere Frage nicht gestellt wurde. Wie wählt man

die Struktur oder Topologie eines HMM, also die Anzahl der Zustände und der zwischen ihnen erlaubten bzw. verbotenen Übergänge?

Die Frage ist aus zweierlei Gründen von Bedeutung. Zum einem ist die Anzahl der zu trainierenden Parameter, vereinfacht gesprochen, quadratisch in der Anzahl der Zustände. Bei einer fest vorgegebenen, beschränkten Datenmenge, wie es in der Praxis fast immer der Fall ist, ergeben sich damit Trainingsfehler für die Parameter, die mit der Modellgröße zunehmen. Des weiteren läuft man mit einer zu großen Anzahl an Parametern Gefahr, das Modell „überzutrainieren“, also Generalisierungsfähigkeit zu verlieren. Zum anderen geht die Anzahl der Zustände auch quadratisch in die Laufzeit der Algorithmen ein, die z.B. für die Datenbanksuche nach verwandten Proteinen benutzt werden.

Sofern das Anwendungsproblem nicht klare Vorgaben lieferte, wurden bisher weitgehend adhoc Verfahren, z.B. die sogenannte Model-Surgery, zur Auswahl einer geeigneten Topologie verwendet. Im Rahmen einer Promotion wurde am ZAIK ein Bayes'scher Ansatz entwickelt, der nach Vorgabe einer Prior-Distribution, die als erwartete Fehlerverteilung der Daten zu interpretieren ist, auf der Basis eines Clusterverfahrens Topologie und Modellparameter lernt.

Mit diesem Verfahren können, ohne Bias für bestehende Problemstellungen, Modelle mit einer geringeren Anzahl an Zuständen und einer größeren Generalisierungsfähigkeit gelernt werden, die darüberhinaus für die typischen Bioinformatik-Anwendungen einen klaren Laufzeitvorteil bieten.

3.2 Analyse von Proteinsequenzen für die Strukturvorhersage von Proteinen

Proteine stehen im Mittelpunkt jedes biologischen Prozesses. Sie katalysieren als Enzyme einen komplexen Ablauf biochemischer Reaktionen, die in ihrer Gesamtheit „das Leben“ ausmachen. Um die molekularen Mechanismen der enzymkatalysierten Reaktionen zu verstehen und beispielsweise in der Medikamentenentwicklung (drug design) in sie eingreifen zu können, ist es notwendig, die 3D-Struktur der beteiligten Proteine zu kennen.

Die experimentelle Bestimmung dieser Struktur ist jedoch wesentlich zeitaufwendiger und teurer als die Bestimmung

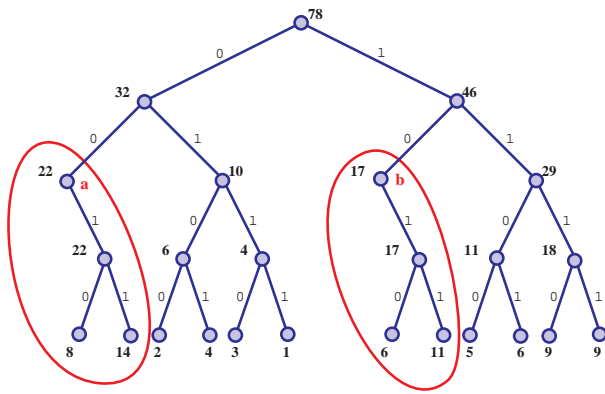


Abbildung 3.2: Sequenzen werden in einen sog. Prefix-Baum eingefügt. Kanten sind mit Ausgabesymbolen gekennzeichnet. Die Knotengewichte entsprechen den Häufigkeiten von Prefixen. Z.B. bedeutet das Knotengewicht 22 des Knotens *a*, dass der Prefix 00 22-mal in den Eingabesequenzen vorkam. In diesem Baum werden die Knoten geclustert, wobei die Cluster dann Zuständen des HMM entsprechen. Die für das Clusterverfahren benutzten Abstände ergeben sich durch Betrachtung der implizit durch die Knotengewichte gegebenen relativen Häufigkeiten (vgl. Wurzel und Blätter der beiden ausgezeichneten Teilbäume).

von Proteinsequenzen: Verdeutlicht wird dies durch die exponentiell wachsende Menge an bekannten Sequenzen im Vergleich zu der wesentlich langsamer wachsenden Anzahl an aufgeklärten Strukturen. Daher ist die Vorhersage der räumlichen Struktur, ausgehend von einer Proteinsequenz, eines der zentralen Probleme der Bioinformatik.

Die Strukturvorhersage funktioniert derzeit allenfalls zufriedenstellend, wenn man die Struktur eines homologen Proteins, d.h. eines Proteins mit gleicher Abstammung, kennt. Dann kann man davon ausgehen, dass eine ähnliche 3D-Struktur vorliegt, und die bekannte Struktur kann als Vorlage (Template) benutzt werden, um mittels Homology Modelling ein Modell für die 3D-Struktur des anderen Proteins zu erstellen.

Im Rahmen einer Kooperation der Arbeitsgruppe Schomburg, Institut für Biochemie, und der Arbeitsgruppe Faigle/Schrader, ZAIK, wird daher versucht, das Auffinden entfernt homologer Proteine zu verbessern.

Das Prinzip des Ansatzes beruht auf der allgemein akzeptierten Annahme, dass zwei Sequenzen mit ausreichend hoher Sequenzähnlichkeit auch eine ähnliche Struktur besitzen. Für die Qualität der Sequenz-Struktur-Zuordnung ist die Sequenzidentität ein entscheidendes Maß. Aber

selbst bei geringer Sequenzähnlichkeit können zwei Proteine homolog sein. Eine Eigenschaft der Homologie ist ihre Transitivität: wenn A und B sowie B und C sich aus dem gleichen Vorfahren ableiten, muss A auch einen Vorfahren mit C gemeinsam haben. Die Transitivität der Homologie wird in diesem Projekt genutzt, um auch entfernt homologe Proteine mit niedriger Sequenzidentität zu finden.

Da Heuristiken zum Sequenzvergleich, wie z.B. Blast und FASTA, bei geringer Sequenzähnlichkeit große Fehler produzieren, haben wir uns entschieden, den rechenaufwendigen Algorithmus nach Smith-Waterman zu verwenden. Für die 86.654 Proteinsequenzen in SwissProt, Release 39, ergibt sich damit ein Gesamtrechenaufwand von über 1.000 CPU-Tagen (UltraSparc CPU). Da das Problem jedoch perfekt zu verteilen ist, konnten wir sämtliche Workstations (ca. 25 CPUs) der Arbeitsgruppe als ein Rechencluster benutzen, und so über einen Zeitraum von 18 Wochen die Berechnung durchführen.

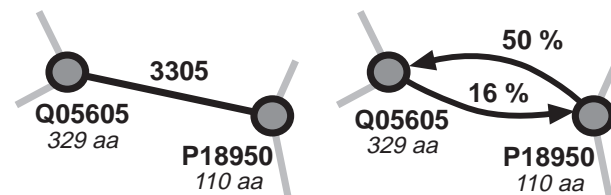


Abbildung 3.3: Vor dem Clustern wird eine ungerichtete Kante, die mit dem Smith-Watermann Alignment-Score gewichtet ist (links) ersetzt durch zwei gerichtete Kanten (rechts). Durch eine Skalierung mit der Sequenzlänge werden unterschiedliche Prozent-Ähnlichkeitswerte auf diesen beiden Kanten erreicht.

Die Ergebnisse der Sequenzvergleiche werden als Eingabe für ein graphenbasiertes Cluster-Verfahren benutzt. Ziel ist dabei herauszufinden, ob sich mittels der resultierenden Cluster tatsächlich mehr entfernt homologe Sequenzen auffinden lassen als bisher. Dies wird durch den Vergleich mit bestehenden Strukturdatensätzen untersucht. Dabei stellen sog. Multidomänenproteine ein Problem dar, dem wir durch Übergang zu einem gerichteten Graphen begegnen konnten.

Die erste Phase dieses Projekts wurde im Rahmen einer gemeinsam von der Arbeitsgruppe Schomburg und Arbeitsgruppe Faigle/Schrader betreuten Diplomarbeit (Eva Bolten, „Eine graphenbasierte Clustermethode zur Detektion entfernt homologer Proteinsequenzen“) durchgeführt. Der Ansatz wird durch die Verwendung der gefundenen Cluster als Trainingsmenge für Profile-basierte Verfahren zum Finden entfernt homologer Sequenzen im Rahmen

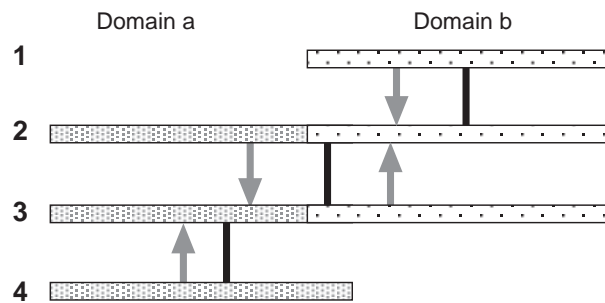


Abbildung 3.4: Das durch Multidomänenproteine entstehende Problem: Auf einem ungerichteten Graphen (schwarze Kanten) sind Proteine #1 und #4 inkorrektweise durch einen Pfad verbunden. Durch den Übergang zu einem gerichteten Graphen (graue Kanten) und einer von der Länge der Proteinsequenzen abhängige Skalierung der Smith-Waterman-Alignmentscores werden solche Kanten vermieden.

einer zweiten gemeinsam betreuten Diplomarbeit ausgebaut. Weiterhin soll der Zugriff auf die Cluster über eine entsprechende Weboberfläche ermöglicht werden.

3.3 Auswahl spezifischer Proben für DNA Arrays

Sowohl Medizin als auch Biologie benötigen effiziente diagnostische Verfahren, um das genetische Erbgut zu analysieren. Die Verfügbarkeit kompletter genomischer Sequenzen ermöglicht es bereits heute, interessante Fragen auf chromosomaler Ebene zu stellen und zu beantworten. Traditionelle Verfahren der Genanalyse wie etwa die Polymerase-Kettenreaktion (PCR), die verschiedenen Blotting-Techniken sowie automatisierte Sequenzierungsverfahren sind jedoch eher sequentieller Natur und für Versuche auf chromosomaler Ebene – bei ca. 3 Milliarden Basenpaare im menschlichen Genom – oder zur schnellen Verarbeitung vieler Proben zu ineffizient und teuer.

Mittels sogenannter DNA-Arrays bzw. DNA-Chips ist es mittlerweile möglich, durch Parallelisierung der Experimente erhebliche Kosten- und Geschwindigkeitsvorteile zu erzielen. Dazu werden einige tausende kurze DNA-Stücke auf vorbestimmten Positionen, sogenannten Spots, eines Glasträgers immobilisiert. Die zu analysierenden (wesentlich längeren) DNA- oder RNA-Sequenzen werden fluoreszent markiert und mit dem Chip zur Reaktion gebracht. Komplementäre Stränge in der zu analysierenden Menge lagern sich dabei auf dem Chip an. Nachdem

überschüssige Polymere durch Waschen des Chips entfernt wurden, kann anhand der Fluoreszenz der einzelnen Spots auf dem Chip bestimmt werden, welche Reaktionen stattgefunden haben. Damit ist es möglich, Informationen über die zu diagnostizierenden Zielsequenzen zu gewinnen.

Das Probe Design Problem

Ein großes Problem beim Design von DNA-Chips ist die Auswahl geeigneter Teilstücke für den Chip. Durch die parallelen Reaktionen auf dem Chip ist es erforderlich, die auf dem Glasträger zu immobilisierenden Stücke so auszuwählen, dass anschließend auch tatsächlich Rückschlüsse auf die Zielsequenzen möglich sind. Dazu sind im wesentlichen drei Bedingungen zu erfüllen:

1. *Spezifität:* Die Proben auf dem Chip sind so zu wählen, dass sie nur mit dem gewünschten komplementären Strang in der Menge der Zielsequenzen binden. Dazu ist für jede Zielsequenz ein kurzes Teilstück auszuwählen, das diese eindeutig charakterisiert, d.h. das in keiner anderen Sequenz vorkommt. Das Problem wird weiter dadurch verkompliziert, dass auch nicht perfekt komplementäre Stränge stabile Bindungen eingehen können. Entsprechende Teilstücke sind keine guten Kandidaten für Proben.
2. *Sensitivität:* Oftmals soll nun mit einem Spot auf dem Chip nicht nur eine einzelne Sequenz erkannt werden, sondern eine ganze Familie verwandter Sequenzen. In diesem Fall muss die Chip-Probe mit jeder Sequenz dieser Familie binden (*muss sensitiv sein*), darf aber mit keiner anderen Sequenz reagieren (*soll spezifisch sein*).
3. *Gleiche Hybridisierungsbedingungen:* Auf der Chipoberfläche müssen einige hundert bis tausend solcher Reaktionen gleichzeitig ablaufen, wobei alle diese Reaktionen unter gleichen Bedingungen, d.h. unter gleichem Salzgehalt in der Reaktionslösung, bei gleichem pH-Wert und vor allem bei gleicher Temperatur stattfinden müssen.

Alle diese Probleme hängen voneinander ab, wodurch ein kompliziertes kombinatorisches Problem entsteht. Im Rahmen einer Diplomarbeit (Lars Kaderali, „Selecting Target Specific Probes for DNA-Arrays“) wurde am ZAIK ein Algorithmus entwickelt, der es dem biologischen Anwender erlaubt, geeignete Kandidaten für das Chipdesign auszuwählen.

Thermodynamisches Modell

Die Stabilität einer Bindung zwischen zwei DNA- bzw. RNA-Stücken kann im Hinblick auf das Chip-Experiment am besten durch die Temperatur ausgedrückt werden, bei der sich die beiden Teilstücke trennen. Sie kann (mittels eines Nearest-Neighbor-Modells) approximativ berechnet werden, unter der Annahme dass

$$T_M = \frac{\Delta H}{\Delta S + R \log C_T}$$

wobei T_M die gesuchte Temperatur, ΔH die Enthalpie-Änderung, ΔS die Entropie-Änderung bei Bindungsreaktion und C_T die Konzentration der DNA bzw. RNA bezeichnet. Damit lässt sich das Probenauswahl-Problem (zunächst für einzelne Sequenzen, unter Vernachlässigung der Familien-Sensitivitäts-Problematik) wie folgt formalisieren: Gegeben n Sequenzen s_1, s_2, \dots, s_n , bestimme eine Temperatur T und n Proben p_1, p_2, \dots, p_n so dass

$$T_M(s_i, p_i) > T > T_M(s_i, p_k) \quad \text{für alle } k \neq i$$

Thermodynamisches Alignment

Das Problem erfordert die Bestimmung der Schmelztemperatur jeder Zielsequenz mit jedem invers-komplementären Teilstück (als Probe). Dabei ist denkbar, dass die beiden nicht linear hybridisieren, sondern einzelne Basen ausgelassen werden und Schleifen oder andere Sekundärstrukturformen bilden. Die Berechnung kann mittels dynamischer Programmierung erfolgen, ähnlich wie in dem bekannten Algorithmus von Needleman-Wunsch bzw. Smith-Waterman, wobei jedoch anstatt der dort üblicherweise verwandten Kostenfunktion hier thermodynamische Parameter verwandt werden. Der Rechenaufwand ist allerdings enorm.

Durch eine Filterung ungeeigneter Teilsequenzen vor Beginn der eigentlichen Rechnungen kann bereits im Vorfeld eine dramatische Reduktion des Rechenaufwandes erreicht werden. Filterkriterien sind z.B. die Eindeutigkeit des gewählten Teilstückes, aber auch die Schmelztemperatur des zugehörigen Duplexes sowie die Länge der Probe. Erst im Anschluss daran beginnt die Berechnung aller möglichen Reaktionen. Abbildung ?? zeigt schematisch den Ablauf dieses Prozesses.

Genomische Sequenzen enthalten viele repetitive Motive. Dadurch kommt es relativ häufig vor, dass Proben gemeinsame Präfixe haben. Da alle Proben gegen die gleichen

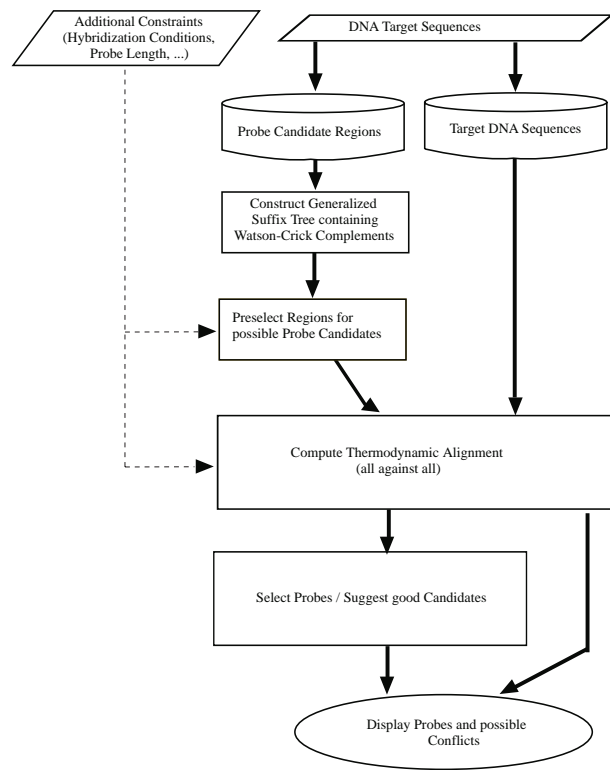


Abbildung 3.5: Schematischer Ablauf des Verfahrens zur Probenauswahl.

Zielsequenzen aligniert werden müssen, kann viel Zeit gespart werden, wenn dann solche Zwischenergebnisse gespeichert werden und bei ihrem erneuten Auftreten nicht neu berechnet werden müssen.

Eine geeignete Datenstruktur zum Erkennen solch gemeinsamer Präfixe sind Suffixbäume. Ein Suffixbaum repräsentiert sehr kompakt alle Teilsequenzen eines Strings. Dabei sind die Kanten des Baumes mit Teilsequenzen beschriftet, und der Weg von der Wurzel zum i -ten Blatt des Baumes liefert genau den i -ten Suffix der ursprünglichen Sequenz. Gemeinsame Präfixe von Suffixen zeichnen sich durch einen gemeinsamen Pfad von der Wurzel aus. Suffixbäume können kanonisch so generalisiert werden, dass sie gleichzeitig die Suffixe mehrerer Sequenzen enthalten.

Die Idee des am ZAIK entwickelten Verfahrens besteht nun darin, einen generalisierten Suffixbaum aus den Zielsequenzen aufzubauen. Dieser wird zunächst anhand oben genannter Kriterien gefiltert, um ungeeignete Proben auszuschließen. Anschliessend wird der Baum in Tiefensuche durchlaufen, und für alle Knoten wird ein Alignment mit den Zielsequenzen berechnet. Dabei müssen Alignments für gemeinsame Präfixe nur einmal berechnet werden.

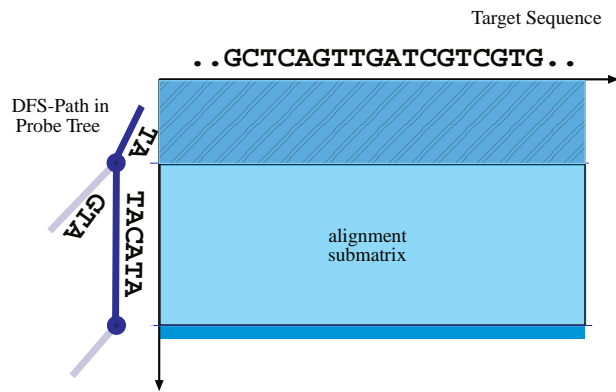


Abbildung 3.6: *Thermodynamisches Alignment mit Hilfe von Bäumen.*

Aus den so berechneten Schmelztemperaturen kann anschließend mit Hilfe von im wesentlichen einfachen Sortierverfahren eine Probenmenge für den Chip ausgewählt werden. Dabei ist es möglich, sowohl im Pre-Processing, d.h. bei der Auswahl von Kandidaten, als auch bei der abschliessenden Auswahl Proben zu bestimmen, die sensitiv für eine ganze Familie von DNA-Sequenzen sind.

Weiterführende Arbeiten

- Die Qualität der durch das Programm bestimmten Proben für tatsächliche Chip-Experimente werden in Zusammenarbeit mit Professor Tautz (Institut für Genetik, Universität zu Köln) bestimmt.
- Sofern keine eindeutigen Proben gefunden werden können, ist es ein einfaches Optimierungsproblem, minimal-konfliktäre Proben zu bestimmen.
- Gegeben eine minimal konfliktäre Probenmenge, muss bei der Chipdaten-Auswertung eine Rückrechnung auf die Daten der einzelnen Proben erfolgen. Dies kann mittels eines statistischen Auswertungsverfahrens, das schon erfolgreich bei statistischen Gruppentests eingesetzt wurde, unter Anwendung von Markov-Chain-Monte-Carlo-Methoden erfolgen.

3.4 Klassifizierung von NMR-Spektren

Die Kernmagnetresonanzspektroskopie (Nuclear Magnetic Resonance, NMR) ist ein sowohl quantitativ als auch qualitativ einsetzbares Verfahren, um Lösungen auf die

in ihnen enthaltenen chemischen Bestandteile zu untersuchen. Durch einen im Vergleich mit konkurrierenden Labormethoden hohen Durchsatz bietet sich der Einsatz der NMR-Spektroskopie für das routinemäßige Untersuchen kompletter Patientenkollektive an.

Wir haben ein Verfahren der statistischen Mustererkennung erfolgreich eingesetzt, um automatisch seltene Stoffwechselkrankheiten bei Neugeborenen zu erkennen. Hierbei werden Proben von Babyurin mit einem NMR-Spektroskop untersucht, und die so erhaltenen Spektren statistisch in Bezug auf die Zugehörigkeit zum Normalkollektiv untersucht. Das Normalkollektiv ist durch einen annotierten Trainingsdatensatz aus ca. 200 Spektren gegeben. Eine Unterscheidung zwischen verschiedenen Krankheiten ist aufgrund des geringen Umfangs des Datenmaterials noch nicht möglich. Unser Ansatz kann aber leicht dahingehend erweitert werden.

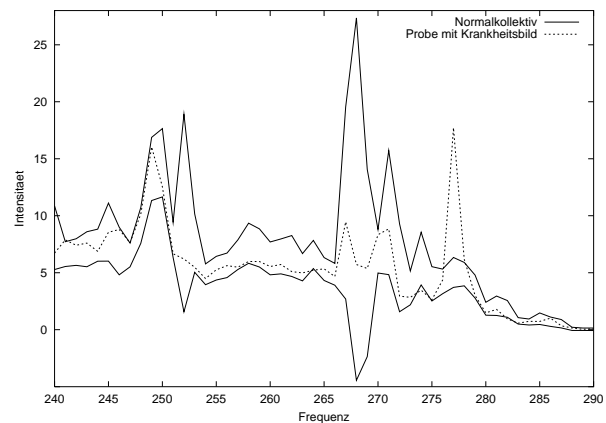


Abbildung 3.7: *Das Normalkollektiv ist durch den Bereich zwischen den durchgezogenen Linien gekennzeichnet. Das Spektrum der Patientenprobe weicht von diesem signifikant ab.*

3.5 Effizientes Screenen von Klon-Bibliotheken

In einer wissenschaftlichen Zusammenarbeit mit der Arbeitsgruppe „Theoretical Biology and Biophysics“ (T-10) am Los Alamos National Laboratory (LANL) in Los Alamos, New Mexico, USA wurden, u.a. im Rahmen mehrerer Forschungsaufenthalte eines ZAIK-Mitarbeiters in Los Alamos, Verfahren entwickelt, um die experimentellen Kosten des Erstellens von physikalischen Kartierungen

gen, einem wichtigen Schritt im Rahmen des Human Genome Projects, zu minimieren.

Gruppentests

Gruppentests lassen sich immer dann erfolgreich einsetzen, wenn eine große Anzahl gleichartiger Objekte demselben Test unterzogen wird. Dabei muss der Test sensitiv genug sein, dann ein positives Ergebnis zu liefern, wenn zumindest ein Objekt der getesteten Gruppe positiv ist. Bei Gruppentests unterscheidet man kombinatorische und statistische Tests. Bei kombinatorischen Tests geht man davon aus, dass das Testergebnis immer korrekt ist, d.h. ein negatives Testergebnis besagt, dass alle Objekte negativ sind, und ein positives Testergebnis besagt, dass zumindest ein Objekt positiv ist. Statistische Tests erlauben den Umgang mit Fehlern und sind damit beim Einsatz in der experimentellen Praxis meist unumgänglich. Gegeben sind hierbei Wahrscheinlichkeiten für das Auftreten falsch negativer Pooltests, also einem negativen Testergebnis trotz positiver Objekte im Pool, und ebenso Wahrscheinlichkeiten für falsch positive Testergebnisse. Häufig hängen diese Wahrscheinlichkeiten von der Anzahl der positiven (negativen) Objekte ab, bei denen man ein falsch negatives (positives) Testergebnis erwarten kann. Je mehr positive Objekte existieren desto unwahrscheinlicher ist ein falsch negativer Pooltest.

Weiterhin unterscheidet man zwischen adaptivem und nicht-adaptivem Gruppentesten: Während bei nicht-adaptiven Gruppentesten die Zuordnung der Objekte zu Pools a priori festgelegt wird, wird beim adaptiven Vorgehen die Zusammenstellung der Gruppen von den Ergebnissen der vorhergehenden Tests abhängig gemacht.

Klon-Bibliotheken und Pooling

Ziel des Humangenomprojektes ist die Entschlüsselung des menschlichen Erbgutes. Ein wichtiger Zwischenschritt ist dabei die Erstellung physikalischer Kartierungen. Um die sehr langen chromosomalen DNA-Moleküle analysieren und sequenzieren – also die Abfolge der Basen A, C, G und T bestimmen – zu können, ist es nötig, eine ausreichende Menge einheitlichen Materials zu gewinnen. Dazu wird die DNA in kleinere Fragmente zerlegt. Chromosomen oder komplette Genome werden mit Restriktionsenzymen in mehrere zehn- oder hunderttausende, einander überlappende Bruchstücke unterteilt. Die Sammlungen solcher Bruchstücke werden als Bibliotheken bezeichnet.

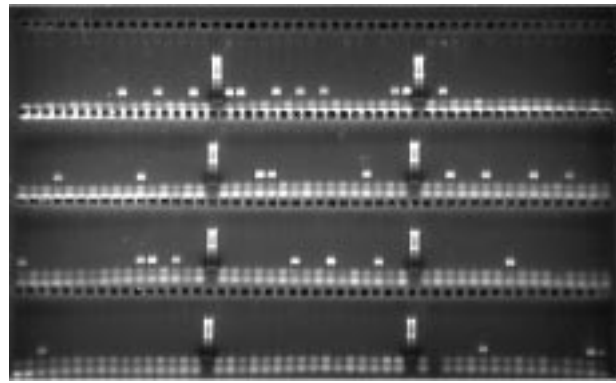


Abbildung 3.8: Das Ergebnis einer PCR-Reaktion wird durch Gel-Elektrophorese sichtbar gemacht. Die hellen Stellen sind genomische DNA.

Einzelne Fragmente werden jeweils z.B. in künstliche He-fechromosomen (YACs, yeast artificial chromosomes) eingebaut, damit man sie vermehren (klonen) kann. Das Zerteilen mit Restriktionsenzymen stellt sich bei unbekannter Sequenz als nicht deterministisch dar: Die Lage der Schnittstellen ist unbekannt. Darüberhinaus ist es nicht sicher, dass an allen möglichen Schnittstellen geschnitten wird. Daher muss, um sicherzustellen, dass die gesamte DNA in der Bibliothek enthalten ist, jede Position in mehreren Klonen enthalten sein. Diese Überdeckungstiefe nennt man Coverage. Typisch sind Werte zwischen 3 und 24.

Nach dem Zerschneiden hat man keine Information über die Lage der in einem Klon enthaltenen DNA in Bezug auf das Genom. Diese muss anhand der Überlappungen der Klone rekonstruiert werden. Hierzu werden sogenannte Sequence-Tagged-Sites (STS) Marker benutzt. Alle Klone, die denselben STS-Marker enthalten, müssen sich überlappen. Zur Überprüfung muss für jede STS und für jeden Klon eine geeignete Polymerase-Chain-Reaktion (PCR) vorgenommen werden. Bei der großen Anzahl an Klonen und zu testenden STSs bieten sich hier Gruppentests an, um den experimentellen Aufwand zu begrenzen.

Um für alle STSs nur einmalig die Pools erstellen zu müssen und damit experimentelle Fehler zu minimieren, haben wir uns für eine nicht-adaptive Poolingstrategie entschieden. Die größte bisher gepoolte Klon-Bibliothek hat 442.368 Klone bei 94 Pools. Hierbei befindet sich jeder Klon in 8 Pools, und jeder der Pools enthält ca. 4.600 Klone.

Statistische Modellierung und Dekodierung

Im Laboratorium wird für jede STS jeweils für alle Pools eine PCR vorgenommen, deren Ergebnisse – enthält einer der Klone im Pool die STS – als „negativ“, „schwach positiv“ oder „stark positiv“ klassifiziert werden. Mit Hilfe eines Bayes'schen Ansatzes können wir ausgehend von den Ergebnissen aller Pools die Wahrscheinlichkeit ausrechnen, mit der ein bestimmter Klon positiv ist.

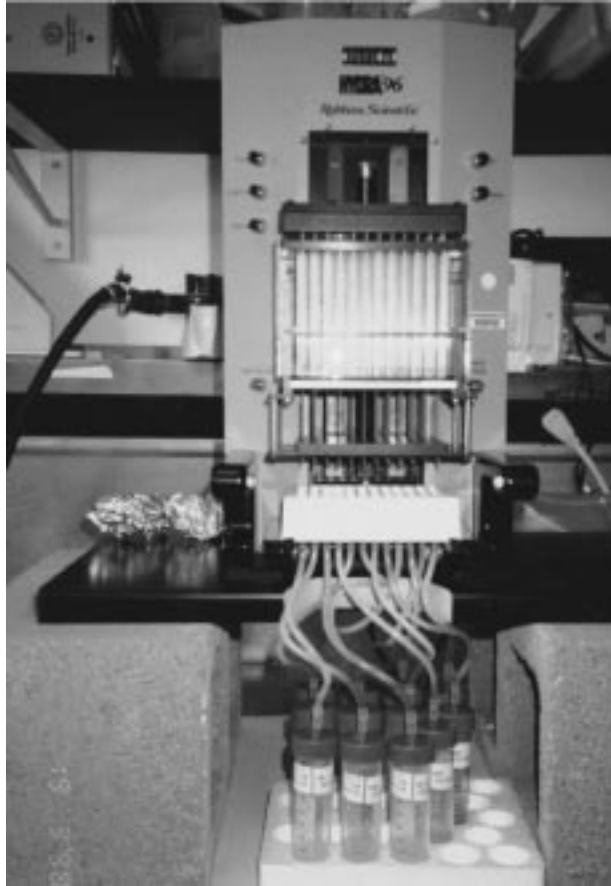


Abbildung 3.9: Mit der Pipettiermaschine *HYDRA* wurden große Teile der Pools erstellt.

Der Ansatz führt allerdings zu einer Summierung über exponentiell viele Terme, was praktisch nicht möglich ist. Um diese Schwierigkeit zu umgehen, benutzen wir eine Markov-Chain-Monte-Carlo-Methode (MCMC). Hierbei bedeutet Monte-Carlo-Methode, dass man sich zufällige Stichproben erzeugt, deren Stichprobenmittelwert z.B. genau die Größe ist, an der man interessiert ist. Bei geeigneten qualitativen Voraussetzungen gibt es Aussagen über die Konvergenz dieses Verfahrens. Bei dem vorliegenden

Problem ist allerdings schon das Erzeugen der zufälligen Stichproben äußerst schwierig. Markov-Ketten liefern in aufeinanderfolgenden Schritten zwar keine zufälligen Stichproben, aber es läßt sich unter bestimmten Bedingungen (z.B. mit Hilfe von Gibbs-Sampling) eine Kette erzeugen, deren Samples für das Mitteln geeignet sind. Dabei garantieren theoretische Resultate die Konvergenz der Kette und damit die Anwendbarkeit des gesamten Verfahrens.

Unser Problem erlaubt eine Berechnung durch die MCMC-Methode, die nicht nur die notwendigen qualitativen Anforderungen erfüllt, sondern mit einer Laufzeit von knapp 30 Minuten (auf einer CPU einer Sun Enterprise 2000) für das Design mit 442.368 Klone auch ausreichend schnell ist.

Ausblick

Unsere Poolingstrategie und die Dekodersoftware laufen seit Anfang 1995 im Produktionsbetrieb im Human Genome Center des Los Alamos National Laboratory. Der Ansatz hat sich im täglichen Einsatz als robust, statistisch zuverlässig und ressourcenschonend erwiesen. Die Performance liegt selbst bei Fehlerraten, jeweils für falsch positive und falsch negative Fehler, von zehn Prozent oberhalb von 85 Prozent.

Bei dem ersten implementierten Poolingdesign war das Verfahren sensitiv genug, Korrelationen zwischen Klonen anzuzeigen, die auf den Mikrotiterplatten benachbart waren. Dies wurde durch weitere Experimente als ein „Überschwappen“ auf den Mikrotiterplatten der Klonbibliothek zurückgeführt, ein Fehler, der ohne das Dekodierungsverfahren vermutlich nie gefunden worden wäre.

Die Auswertungstrategie lässt sich auch auf Problemstellungen bei DNA-Chips anwenden, und zwar wenn keine eindeutigen sequenzspezifischen Proben gefunden werden können. In dieser Hinsicht laufen Untersuchungen in Zusammenhang mit dem Projekt unserer Arbeitsgruppe zur Selektion von Proben für DNA-Chips.

Dies ist eine gemeinsame Arbeit mit David C. Torney (Group T-10, LANL). Weiterhin beteiligt sind Manny Knill, Norman Dogget, Jon. A. Longmire, Judy Tesmer, David Bruce, Bill Bruno (LANL).

3.6 Modellierung metabolischer Netzwerke

Aufgrund der stetig anwachsenden Menge an genetischer Information durch die Entschlüsselung kompletter

Genome entsteht in vielen Bereichen der Biologie eine Vielfalt neuer Fragestellungen. Hierbei betreffen viele offene Probleme den Metabolismus einer Zelle. Unter dem Metabolismus versteht man die Gesamtheit der chemischen Reaktionen, die für die Bereitstellung der freien Energie verantwortlich sind, damit für Erhaltung, Teilung etc. sorgen und somit wiederum die Ausübung ihrer Funktionen sichern.

Nahezu alle metabolischen Reaktionen sind enzymatische Reaktionen und werden so durch zelleigene Proteine gesteuert, womit der Zusammenhang Metabolismus-Genom hergestellt ist. Wichtige Subsysteme des Metabolismus sind sogenannte metabolische Pfade. Ein metabolischer Pfad ist eine Abfolge von Reaktionen, so daß jeweils das Endprodukt der vorhergehenden Reaktion das Ausgangsprodukt der nächsten ist (inklusive Verzweigungen oder auch Kreisen). Auf diesen Wegen werden zellwichtige Zwischen- und Endprodukte synthetisiert.

Im Rahmen der Promotion eines ZAIK-Mitarbeiters werden neue Verfahren zum Schaffen von "Substraträumen" für enzymatische Reaktionen entwickelt. Diese Substraträume enthalten eine möglichst umfassende Auswahl von Substanzen, die von einem Enzym prozessiert werden bzw. das Enzym blockieren und somit die enzymatische Reaktion inhibieren. Wir versehen unsere Substraträume mit einem geeigneten Abstandsmaß und bekommen so mit Hilfe bereits gut bekannter Substrate ein Maß dafür zurück, wie wahrscheinlich noch unerforschte Stoffe von einem Enzym prozessiert werden bzw. das Enzym inhibieren. Auf ähnlichem Wege schaffen wir "Produkt-Räume", die eine möglichst umfassende Auswahl von Substanzen enthalten, die als Produkt einer enzymatischen Reaktion in Frage kommen. Durch Iterieren, d.h. durch Bilden von Durchschnitten von Substrat- und Produkträumen lassen sich ganze Kaskaden enzymatischer Reaktionen durchspielen. Damit lassen sich die Mechanismen noch unerforschter Substanzen wie z.B. neuer Medikamente simulieren bzw. deren Risiken abschätzen.

gegründet. BIG steht allen aus der Region Köln-Bonn offen, die ein Interesse an Methoden, Verfahren, Problemen und Aufgabenstellungen der Bioinformatik haben, die sich mit Gleichgesinnten austauschen möchten, oder einfach nur einen ersten Einblick in das Gebiet bekommen möchten. Dabei richtet sich BIG sowohl an Studenten und Mitarbeiter universitärer Institute als auch an Angestellte kommerzieller Unternehmen, um einen möglichst praxisnahen und engen Austausch zwischen Bildung, Forschung und Industrie zu fördern.

In regelmäßigem Abstand veranstaltet BIG dazu Treffen, die jeweils einem aktuellen Thema gewidmet sind. So wurden bei vergangenen Treffen unter anderem DNA-Chips, Homologien bei Proteinsequenzen oder Data Mining Probleme vorgestellt und diskutiert.

Um den Vernetzungsgedanken weiter zu unterstützen, wurden ausserdem zwei öffentliche Mailinglisten eingerichtet. Über *big-announce@zpr.uni-koeln.de* können Termine zu Veranstaltungen zum Thema Bioinformatik bekanntgegeben werden. Über diese Mailingliste werden auch die regelmässigen BIG-Treffen angekündigt. Über die Liste *big-discussion@zpr.uni-koeln.de* können inhaltliche Diskussionen geführt und Fragen erörtert werden, die sich auf Verfahren, Probleme, Tools und Aufgabenstellungen der Bioinformatik beziehen oder sonst von allgemeinem Interesse sind.

Weitere Informationen über BIG und über die Mailinglisten (und wie man sich für diese anmeldet) können im Internet unter <http://www.zaik.uni-koeln.de/big> abgerufen werden, oder Kontaktaufnahme per E-Mail an big@zpr.uni-koeln.de.

Kontakt: bioinformatik@zpr.uni-koeln.de

3.7 Bioinformatik

Interessengruppe BIG

Zur Vernetzung und Förderung des wissenschaftlichen Austausches in der Bioinformatik wurde im vergangenen Jahr die Bioinformatik-Interessen-Gruppe ("BIG")

Kapitel 4

Umweltmanagement

Seit der Konferenz 1992 in Rio de Janeiro, auf dem die erste Klimaschutzrahmenkonvention verabschiedet wurde, sehen es die unterzeichnenden Staaten als gemeinsame Aufgabe an, die Klimaproblematik wissenschaftlich zu untersuchen und den möglichen Umweltveränderungen zu begegnen. In dem Vertragstext wurde ein kooperatives Vorgehen innerhalb von internationalen Projekten als mögliche Handlungsoption explizit formuliert. Ein zentrales Problem bei derartigen Unternehmungen stellt die vertragliche Ausgestaltung der Technologiekooperationen dar. Darüber hinaus kommt der Aufteilung der entstehenden Kosten eine besondere Bedeutung zu.

Auf der 1. Vertragsstaatenkonferenz 1995 zum Rahmenübereinkommen über Klimaveränderungen in Berlin wurde daraufhin beschlossen, Joint-Implementation Programme in einem verstärkten Maße einzusetzen. Joint-Implementation meint hierbei die gemeinsame Umsetzung von Klimaschutzzielen im Rahmen von Technologiekooperationen. Im Gegensatz zu anderen Instrumentarien zur CO_2 -Minderung trägt Joint-Implementation einen kooperativen Charakter.

Da bisher nur geringe Projekterfahrung mit Joint-Implementation vorliegt, ist man auf eine Simulation von verschiedenen Handlungsszenarien angewiesen. Dies legt eine numerische Analyse nahe. Gleichzeitig kann durch eine kontrolltheoretische Behandlung des Problemzusammenhanges versucht werden, die Koordination von Handlungsaktivitäten über einen längeren Zeitraum im Rahmen eines Umweltmanagements gezielt zu fördern. Es ist daher der Schwerpunkt des mit dem Titel beschriebenen Forschungsvorhabens, ein Joint-Implementation Programm zu modellieren und mit einer mathematischen Theorie derart auszugestalten, dass eine qualitative Beurteilung dieses neuen Instrumentariums möglich wird. Um eine solche Analyse auf komfortable Weise zu ermöglichen, wurde das TEM-Modell (Technologie-Emissionen-Mittel-) Modell entwickelt und analytisch untersucht.

4.1 Modellierung und Simulation von Emissions-Minderungsaktivitäten am Beispiel von Joint-Implementation unter Verwendung des TEM-Modells

Das TEM-Modell basiert auf dem allgemeinen SCX-Konfliktmodell, das von Jürgen Scheffran an der Universität Marburg entwickelt wurde. Für das TEM-Modell wurde ein Kostenspiel eingeführt und der definierbare Core als zulässige Steuerungsmenge eines Kontrollproblems betrachtet. Auf diese Weise konnte innovativ die kooperative Spieltheorie mit der Theorie zeitdiskreter Systeme verbunden werden. Die kombinatorische Struktur der zulässigen Mengen ermöglicht es, Aussagen über die Eigenschaften der Lösungstrajektorien zu treffen.

Im Zentrum des Interesses steht die Frage nach der möglichen Reduzierung von existierenden CO_2 -Emissionen bzw. die Erreichung von bestimmten Klimaschutzzielen. Diese wurden auf globaler Ebene im Rahmen von Klimaschutzvereinbarungen getroffen.

Besondere Bedeutung kommt in diesem Zusammenhang Projekten zwischen der Volksrepublik China und einem Staat der Europäischen Union zu. Die Kraftwerke der Volksrepublik China besitzen derzeit einen durchschnittlichen Nettowirkungsgrad von 22%, der um ca. 40% unter dem Durchschnitt liegt, der von bundesdeutschen Kraftwerken erreicht wird. Gleichzeitig wird China bei einer Fortsetzung der momentanen Entwicklung in den nächsten 15 Jahren die USA als den weltweit größten Emittenten an CO_2 übertreffen. So müßte die europäische Union allein ihren Beitrag an den globalen CO_2 -Emissionen im Jahr

1990 um nahezu 85% bis zum Jahre 2010 senken, um den CO_2 -Anstieg allein in China kompensieren zu können. Eine Reduktion für die einzelnen Akteure kann nun durch folgende Maßnahmen erreicht werden:

- Finanzielle Mittelaufwendungen
- Effizienzsteigerungen
- Innovationen
- Technologiekooperationen

Das TEM-Modell zeichnet sich insbesondere dadurch aus, dass die gleichzeitige Berücksichtigung dieser Möglichkeiten in einem einzigen Modellansatz integriert ist. Ferner wird in der zugrunde liegenden Dissertation ausführlich die Möglichkeit einer Fondsbildung behandelt.

4.2 Grundgleichungen des TEM-Modells

Im TEM-Modell beschreiben die Parameter E die Reduktion der Emissionen und M die finanziellen Mittelaufwendungen der einzelnen Akteure. Das Modell umfaßt n Akteure. Dies wird durch den Index i angezeigt. Die zweite Gleichung besitzt einen Proportionalitätsfaktor λ , der das Maß der Investitionen beschreibt, sowie einen Gedächtnisparameter φ , der die logistische Differenzengleichung um einen Verzögerungsfaktor erweitert. Die Variable M^* stellt jeweils eine obere Budgetgrenze für die einzelnen Akteure dar.

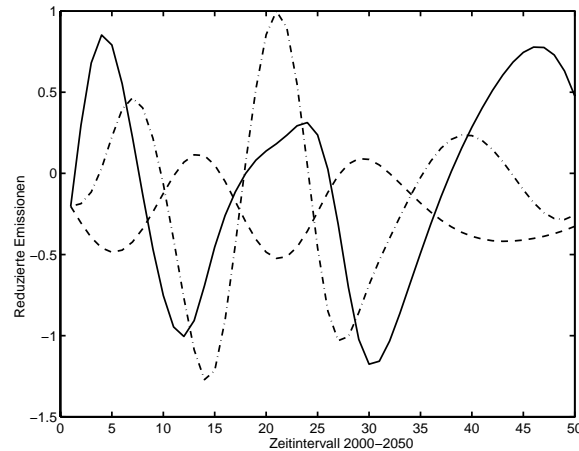
$$\begin{aligned}\Delta E_i(t) &= \sum_{j=1}^n em_{ij}(t)M_j(t) \\ \Delta M_i(t) &= -\lambda_i M_i(t)[M_i^* - M_i(t)][E_i(t) + \varphi_i \Delta_i E_i(t)]\end{aligned}$$

$E_i(t)$	Reduzierte Emissionen
$M_i(t)$	Finanzielle Mittelaufwendungen
M_i^*	Budgetgrenze
λ_i	Wachstumsparameter
φ_i	Gedächtnisparameter

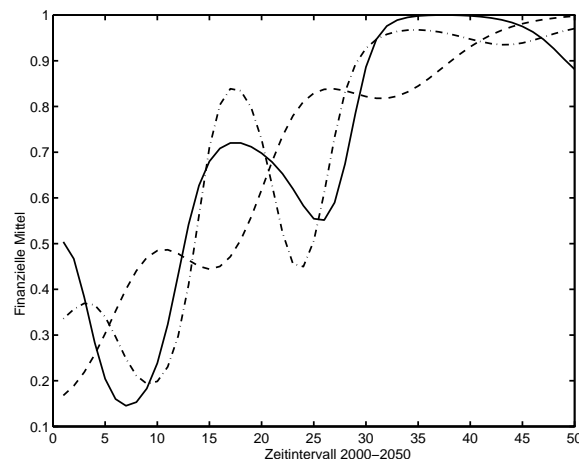
Im Zentrum des TEM-Modells stehen die em -Parameter, die eine Art Effektivitätsmatrix darstellen und die Koppelung der beiden nichtlinearen Gleichungen bewirken.

4.3 Simulation

Simulationen machen deutlich, dass die nichtlineare Struktur kein gemeinsames Erreichen des gewünschten Endzustands garantiert. Die Akteure starten in dem normalisierten Szenario unterhalb der Koordinatenachse („Kyoto-Linie“) im negativen Bereich. Die Emissionskurven weisen für das in der angegebenen Dissertation ausführlich beschriebene Beispiel zudem ein starkes Oszillieren auf.



Obwohl kein Akteur den gewünschten Zielzustand dauerhaft erreicht, ist bei allen Akteuren ein Anwachsen der finanziellen Mittelaufwendungen zu beobachten.



Dies legt die Entwicklung eines steuernd wirkenden Umweltmanagementsystems nahe.

4.4 Umweltlizenzen

Seit mehr als 30 Jahren werden innerhalb der Umweltökonomie Umweltlizenzen als Alternative zu existierenden ordnungsrechtlichen Verfahren angesehen. In dem Kyoto Protokoll (siehe hierzu auch: Oberthür, S; Ott, H.E., Das Kyoto Protokoll - International Climate Policy for the 21st Century, International and European Environmental Policy Series, Springer Verlag Berlin-Heidelberg 2000) werden handelbare Umweltrechte sogar explizit als marktwirtschaftliche Instrumente empfohlen. Damit kann über die ökonomischen Steuerungsanreize ein effizienter Beitrag zum Umweltschutz geleistet werden. Für die Entwicklung und Etablierung eines solchen Umweltlizenzensystems werden nach Hansjürgens (Handelbare Umweltlizenzen. In: ZiF Jahresbericht 1998/99, Universität Bielefeld) die folgenden vier Phasen als notwendig angesehen:

1. Festlegung einer zulässigen Gesamtemissionsmenge
2. Verteilung der Lizenzen, z.B. durch ein *grandfathering*-Verfahren
3. Vorkehrungen für einen funktionierenden Handel
4. Entwicklung von Kontrollmechanismen

Daher kann die Reduktion von CO_2 -Emissionen als exemplarisch für die Etablierung eines solchen Umweltlizenzensystems angesehen werden. In diesem Zusammenhang kommt der Bewertung von verschiedenen Technologiepfaden eine zentrale Bedeutung zu.

4.5 Umweltsystemforschung

Das TEM-Modell bietet die Möglichkeit, die Umsetzung verschiedener Technologiepfade innerhalb eines Energiemanagements über einen längeren Zeitraum zu beschreiben und die damit verbundenen Auswirkungen auf CO_2 -Minderungsaktivitäten zu beurteilen. Desweiteren werden durch die Berechnung optimaler zulässiger Trajektorien Handlungskorridore bestimmt, die sich für die konkrete Umsetzung im Rahmen eines verantwortungsvollen und nachhaltigen Umweltmanagements anbieten.

Es ist das Ziel zukünftiger Forschungsvorhaben, die gewonnenen Ergebnisse aktuellen Fallstudien gegenüberzustellen. Hierbei erweist es sich von großem Vorteil, dass bei der Entwicklung des TEM-Modells nur Größen berücksichtigt wurden, die empirisch bestimmbar sind.

An der Technischen Universität Darmstadt wird in der IANUS-Gruppe (Interdisziplinäre Arbeitsgruppe Naturwissenschaft, Technik und Sicherheit) in einem gemeinsamen Forschungsprojekt mit dem ZIT (Zentrum für Interdisziplinäre Technikforschung) - Projektpartner sind Prof. Dr. W. Krabs (Fachbereich Mathematik) und Prof. Dr. D. Ipsen (Institut für Volkswirtschaftslehre), beide TU Darmstadt - an einer Fallstudie zu Joint-Implementation gearbeitet, in die die beschriebenen Ergebnisse einfließen werden. Dadurch wird eine Umweltsystemforschung angestrebt, die sich auf reale ökonomische Zusammenhänge stützt.

In einem weiteren Schritt ist es vorgesehen, das TEM-Modell hinsichtlich der Simulation eines internationalen Zertifikatenhandels (Joint International Emissions Trading - IET) auszuweiten.

Es besteht ein enger Forschungskontakt zum Umweltforum der Universität zu Köln sowie zu dem Potsdamer Institut für Klimafolgenforschung (PIK).

Das Forschungsprojekt dient insbesondere der Weiterentwicklung des TEM-Modells als Umweltmanagementsystem, das im Jahr 2000 mit dem Dissertationspreis der Gesellschaft für Operations Research GOR gefördert wurde.

Kontakt: umweltmanagement@zpr.uni-koeln.de

Kapitel 5

Multimedia

Das ZAIK ist in mehreren Projekten an der Gestaltung und Weiterentwicklung von multimedialen Lehr- und Lerneinheiten beteiligt. Im Rahmen des europäischen Projektes „Trial-Solution“ wird die sogenannte Slicing-Book-Technologie entwickelt und erprobt, die es ermöglicht, Lehr- und Lernmaterialien individuell zu modularisieren und multimedial aufzubereiten.

Zu den ausgewählten Materialien wird die Combinatorial Algorithm Toolbox (CATBox), die 2001 im Springer Verlag erscheinen wird, zählen. Bei ihr handelt es sich um einen interaktiven Kurs in diskreter Mathematik, der am ZAIK entwickelt wurde. CATBox basiert auf der Graphenvisualisierungssoftware Gato (Graph Animation Toolbox).

Im Mittelpunkt des NIKOS-Projektes, das gemeinsam mit dem Seminar für Politikwissenschaft durchgeführt wurde, stand die Entwicklung eines virtuellen ONLINE-Seminars für Studierende aus dem beteiligten Fachgebiet. Die computergestützte interaktive Simulation von politischen Handlungszusammenhängen in einem virtuellen Entscheidungsraum konnte insbesondere das selbstorganisierte Lernen von komplexen Entscheidungsprozessen besonders fördern.

Kontakt: multimedia@zpr.uni-koeln.de

5.1 Trial Solution

Unter dem Titel „*Trial-Solution*“ firmiert ein Forschungsprojekt, an dem neben dem ZAIK fünf Hochschulen, zwei Verlage und vier Unternehmen und Institutionen aus fünf europäischen Ländern beteiligt sind. Die Europäische Union unterstützt das Forschungsvorhaben innerhalb der nächsten drei Jahre mit 1,5 Mio. Euro. Durch dieses Projekt soll die internetbasierte Slicing Book Technologie, die Technologie des modularisierten Buches, weiterentwickelt werden.



Tools for
Reusable
Integrated
Adaptable
Learning Systems
Standards for
Open
Learning
Using
Tested
Interoperable
Objects
Networking

Dieses Verfahren ist von Dr. Bernd Ingo Dahn, der Projektleiter dieses europaweiten Projektes ist, in der Arbeitsgruppe „Künstliche Intelligenz“ von Prof. Dr. Ulrich Furbach des Koblenzer Fachbereichs Informatik als Beitrag zu einer Symbiose von Buch und Internet entwickelt worden.

5.1.1 Die Slicing Book Technologie

Die Grundidee hinter der Slicing Book Technologie ist einfach zu beschreiben: Man zerlege ein Buch in kleine Teile und setze daraus neue Bücher zusammen. Durch diese Form der Modularisierung können die Inhalte an die individuellen Bedürfnisse angepasst werden. Die Leser werden durch ein Beratungssystem unterstützt, das sowohl inhaltliche Zusammenhänge als auch Vorkenntnisse der Leser berücksichtigt. Die durch das Zerlegen entstandenen sogenannten semantischen Einheiten können Kapitel, Absätze, Abbildungen oder Tabellen und Grafiken sein. Ihre Verwaltung erfolgt über Datenbanken. Besonders geeignet ist die Slicing Book Technologie für Lehrbücher und juristische oder technische Dokumente. Dozenten bietet sich dadurch die Möglichkeit, individuelle Lerneinheiten zu gestalten, die auf spezifische Lernsituationen und Lernziele hin entwickelt werden können.

5.1.2 Individuelle Modularisierung und Evaluation

Im Rahmen des EU-geförderten Projektes Trial Solution wird eine Bibliothek von etwa 20 aufbereiteten Lehrmaterialien für Fächer wie Mathematik, Informatik, Ingenieurwissenschaften oder Wirtschaftswissenschaften entwickelt. Diese Materialien werden zusätzlich in der Lehre an den beteiligten Hochschulen erprobt. Dadurch soll die Wiederverwendung hochwertiger Lehrmaterialien bei unterschiedlichen Lehrbedingungen unterstützt werden. Bisher ist ein Mathematiklehrbuch mit Hilfe dieser Technologie aufbereitet worden, das unter dem Namen *Analysis Individuell* (Wolter, H.; Dahn, B.I.: Analysis Individuell. Springer Verlag, Berlin 2000) erschienen ist.

Das Lehrbuch *Analysis Individuell* besteht aus drei Komponenten: Die gedruckte Version beinhaltet ein Repetitorium zur Grundvorlesung Analysis. Es umfasst in strukturierter und sich auf das Wesentliche beschränkende Form den zugrundeliegenden Stoff sowie das Basiswissen für Vordiplom und Zwischenprüfungen. Die beigelegte CD-ROM enthält das vollständige Lehrbuch mit den ausführlichen Beweisen. Mit dem Kauf der CD-ROM ist der Erwerb einer Zugangsberechtigung verbunden, die es ermöglicht, via Internet das auf einem Server liegende Buch nach *individuellen* Wissensbedürfnissen zu konfigurieren. Diese auf die speziellen Bedürfnisse abgestimmte Vorgehensweise ist derzeit einzigartig.

Zu den weiteren ausgewählten Materialien wird CATBox (Combinatorial Algorithm Toolbox) zählen. Hierbei handelt es sich um einen interaktiven Kurs in Diskreter Mathematik, der am ZAIK entwickelt wurde und im Springer Verlag erscheinen wird. Die Modularisierung und Evaluation von CATBox wird am ZAIK durchgeführt.

5.1.3 Multimediaeinsatz und Inter-Buch-Verweise

Die Projektpartner werden insbesondere das Verbinden der verschiedenen Lehrmaterialien erproben. Dieses Verknüpfen bietet den Vorteil, dass ein Problemzusammenhang nicht mehrfach behandelt werden muss, sondern vielmehr zu neuen Einheiten zusammengestellt werden kann. Diese Einheiten können dann leicht auf spezifische Lehr- und Lernsituationen abgestimmt und optimiert werden. Die Verknüpfung kann über sogenannte Inter-Buch-Verweise erfolgen. Es kann jedoch auf verschiedene Weisen dargestellt oder aus verschiedenen Blickwinkeln betrachtet werden. Mit Hilfe der Verweise können zusätzlich verwandte Beispiele aus anderen thematischen Zusammenhängen leicht gefunden werden. Die Slicing Book Technologie bietet auch in der Form der Darstellung keine Einschränkungen. Texte, Tabellen oder Grafiken können ebenso umgesetzt werden wie alle Präsentationsformen, die im Internet verwendet werden können wie z.B. Video-clips oder Animationen. Somit bestehen größere Möglichkeiten als beispielsweise bei den sogenannten E-Books.

5.1.4 Aufbau einer virtuellen Bibliothek von Lernmaterialien

Es ist das Ziel des Projektes, verschiedene Möglichkeiten des Multimediaeinsatzes miteinander zu kombinieren und dadurch eine virtuelle Bibliothek von unterschiedlichen Lernmaterialien anzulegen. Ein weiterer Vorteil dieser Form des elektronischen Publizierens sind die vielfältigen Suchmöglichkeiten. Neben der Volltextsuche ist die Suche nach Schlüsselbegriffen möglich. Selbst wenn ein Schlüsselbegriff nicht genau bekannt ist, kann die Recherche über Synonyme oder Unterbegriffe fortgesetzt werden.

Kontakt: trialsolution@zpr.un-koeln.de

5.2 Graph Animation Toolbox

Gato, die Graph Animation Toolbox, ist ein multimediales Werkzeug für die Visualisierung von Graphenal-

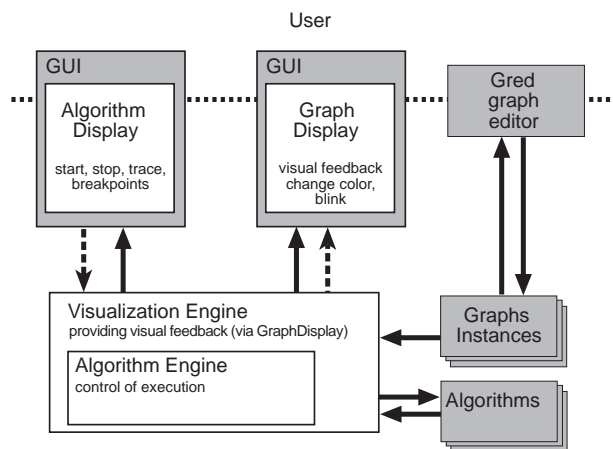


Abbildung 5.1: Die einzelnen Funktionseinheiten in Gato: Gestrichelte Linien bezeichnen eine Kontrollbeziehung und gefüllte Linien Datenfluß.

gorithmen. Gato erlaubt es Studenten und Wissenschaftlern, sowohl mit Algorithmen als auch mit Graphen zu experimentieren. Gato benutzt eine graphische Benutzeroberfläche und läuft unter Linux, Unix, Microsoft Windows 95/98/NT und unter Macintosh OS. Um die Einarbeitungszeit zu minimieren, entsprechen die Elemente der Benutzeroberfläche dem jeweiligen Systemstandard (native look-and-feel). Außerdem ist die Bedienung an bekannte Konzepte („Debugger“, Zeichenprogramm) angelehnt.

Die Darstellung der Algorithmen ist zweigeteilt: Die Anweisungen werden im *Algorithmusfenster* in der Programmiersprache Python, die dem aus anderen Lehrbüchern bekannten Pseudocode ähnelt, angezeigt. Im *Graphfenster* wird die Problemistanz, also der Graph, auf dem der Algorithmus abläuft, und die Arbeitsweise des Algorithmus detailliert visualisiert.

Der Ablauf kann vom Benutzer mittels der von „Debuggern“ bekannten Eingriffsmöglichkeiten gesteuert werden. Ein Algorithmus kann vollständig oder zeilenweise ablaufen. Die Ausführung kann an sogenannten, vom Benutzer zu definierenden „Breakpoints“ unterbrochen werden. Darüber hinaus bieten sich dem Benutzer weitere Möglichkeiten, aktiv in den Ablauf einzugreifen.

Die Auswirkungen der Anweisungen in dieser Zeile auf den Graphen werden durch Änderungen von Kanten- und Knoten-Farben, Blinken und weitere optische Signale im Graphfenster unmittelbar verdeutlicht. Hierdurch lassen sich die Zusammenhänge zwischen den Vorschriften des Algorithmus und den Auswirkungen auf die jeweilige Problemistanz direkt und schnell erfassen. Desweiteren ist

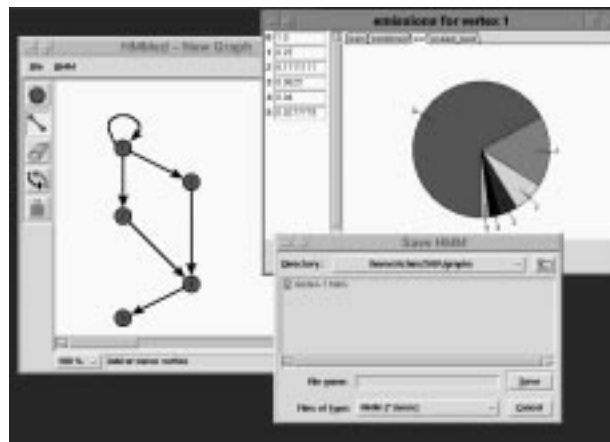


Abbildung 5.2: Ein experimenteller Editor für Hidden-Markov-Modelle, der auf Gato basiert.

Gred, ein einfacher Grapheditor zur interaktiven Erstellung und Veränderung von Graphen, integriert. Für die Editierung der Algorithmen kann ein beliebiger Texteditor benutzt werden.

Für das Erstellen der Visualisierung ist ein regel-basiertes Konzept in Form von Animated-Data-Structures (ADS) umgesetzt worden. Durch diesen Ansatz lässt sich bei geringem Aufwand ein hohes Maß an Konsistenz erreichen. Durch den Einsatz der ADS ergibt sich weiterhin eine weitgehend „automatische“ Visualisierung, die es erlaubt, bestehende Algorithmen unter Beibehaltung der Visualisierung zu ändern bzw. neue Algorithmen mit geringem Aufwand visualisiert zu implementieren.

Anwendungen

Gato ist unter Library Gnu Public License (LGPL) veröffentlicht. Diese Lizenzierung hat Vorteile in Bezug auf die Einsetzbarkeit von Gato. Neben der im nächsten Abschnitt beschriebenen CATBox, wird Gato bereits für „AsIF – an authoring System for interactive fiction“ außerhalb der Arbeitsgruppe benutzt.

Im Zusammenhang mit dem Einsatz von Hidden-Markov-Modellen (HMM) in der Bioinformatik und im Bausparkassenprojekt und einer Integration mit der in der Arbeitsgruppe entwickelten HMM-Softwarebibliothek wird auf der Basis von Gato in einem ersten Schritt ein graphischer Editor für HMM entwickelt. Die Integration der HMM-Softwarebibliothek und die Schaffung von Optionen zum „Computational Steering“ ist das Folgeziel.

Kontakt: gato@zpr.uni-koeln.de

CATBox

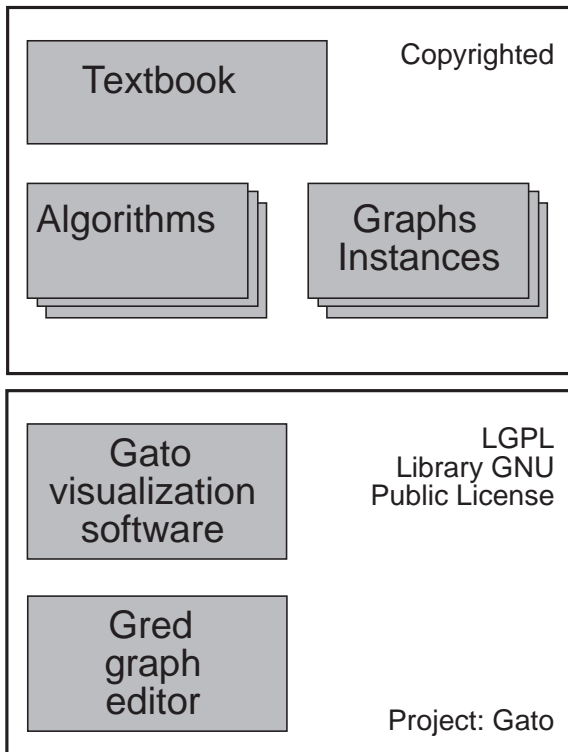


Abbildung 5.3: Überblick über die einzelnen Komponenten im CATBox-Projekt.

5.3 CATBox

CATBox (**C**ombinatorial **A**lgorithm **T**ool**B**ox) ist ein in sich abgeschlossenes Lernsystem zur Einführung in die algorithmische Kombinatorik, das ein Lehrbuch und die Visualisierungs-Software Gato kombiniert. Im Buch werden kombinatorische Probleme anhand praktischer Anwendungen motiviert. Zugrundeliegende mathematische Ideen und Hilfsmittel werden ausführlich dargelegt, und es wird auf die Modellierung und auf Lösungsansätze eingegangen. Die Software visualisiert die Algorithmen und erleichtert so einerseits das Verständnis und erlaubt andererseits eine Vertiefung des Wissens durch Experimente mit Graphen und Algorithmen.

Das Ziel von CATBox ist die Vermittlung kombinatorischer Ideen anhand ausgewählter Problemstellungen. Die Erfahrungen mit dem seit 1987 bestehenden Vorgängersystemen und vorhandenen Lehrbüchern haben gezeigt, daß weder ein Programm noch ein Lehrbuch alleine einen optimalen Zugang bieten kann. Die Software ermöglicht es, die Dynamik eines Algorithmus darzustellen. Ein Buch ist

ein geeigneteres Medium für die Vermittlung der mathematischen Grundlagen und Analysen von Algorithmen.

Darüber hinaus macht erst die Kombination aus Lehrbuch und Software visuell unterstütztes Lernen durch Interaktion mit dem Programm möglich. Der Benutzer wird so zum Betrachten, Experimentieren und Mitmachen eingeladen. Zielgruppe sind dabei die Hörer von Hochschulvorlesungen der Mathematik und Informatik.

Aufbau des Buches

In jedem Kapitel wird eine Problemklasse durch ein Beispielproblem vorgestellt. Die Praxisrelevanz der Problemklasse wird durch eine Reihe von Anwendungen motiviert. Im Anschluß werden Überlegungen vorgestellt, die zur Lösung des Problems geeignet scheinen und deren Erfolg bzw. Mißerfolg anhand der vorgestellten Probleme erläutert und begründet. Erfolgreiche Ideen werden in Algorithmen umgesetzt. Die Funktionsweise und Komplexität der Algorithmen wird im Programm verdeutlicht und dient als Motivation für den Beweis der Operationalität und die Diskussion der Komplexität im Buch.

Das Buch behandelt fünf Problemklassen. Zunächst wird das Problem, minimale aufspannende Bäume in Graphen zu finden, und seine erfolgreiche Lösung vorgestellt. Weitere Kapitel widmen sich dem Kürzeste-Wege-Problem, dem Maximaler-Fluß-Problem und dem Kardinalitätsmatching. Abschließend werden primal-duale Algorithmen zur Lösung gewichteter Matching-Probleme diskutiert.

Für die im Lehrbuch vorgestellten Algorithmen liegt eine Bibliothek von Beispielgraphen vor. Die Algorithmen zeigen auf den jeweiligen Beispielgraphen sowohl ihr typisches, aber auch ihr „pathologisches“ Verhalten. Der in Gato integrierten Grapheneditor erlaubt es dem Benutzer z.B. zu beobachten, inwieweit sich Änderungen der Beispielgraphen auf das Verhalten der Algorithmen auswirken.

CATBox wurde im Rahmen von Vorlesungen unserer Arbeitsgruppe Faigle/Schrader an der bzw. an der TU Cottbus erfolgreich eingesetzt. Buch und Software werden im Herbst 2001 im Springer-Verlag erscheinen.

Kontakt: catbox@zpr.uni-koeln.de

5.4 Nikos

Im Zentrum des Projektes „Neue Informations- und Kommunikationstechnologien im Studium“ (NIKOS) steht ein

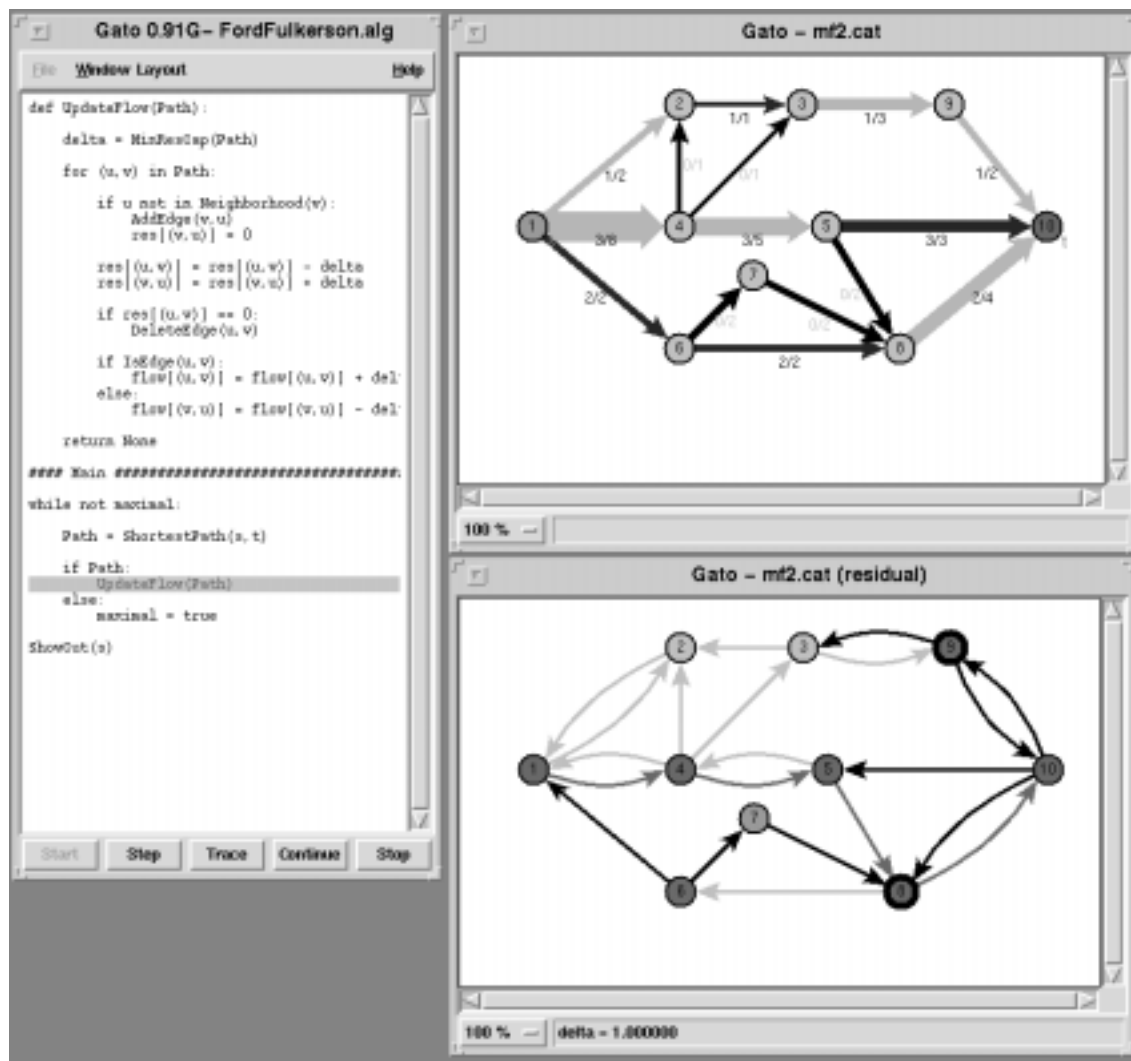


Abbildung 5.4: Hier wird die Visualisierung des Maximalen-Fluß-Algorithmus nach Ford-Fulkerson in Gato dargestellt. Das Algorithmfenster ist links gezeigt, das Graphenfenster (oben) und das Restnetzwerk (unten) auf der rechten Seite.

neuartiges multimediales Seminar, das den Teilnehmern WWW-basierten Zugang zu einer umfassenden Simulation ökonomischer, politischer und sozialer Vorgänge in globaler Perspektive bietet: Das NIKOS-Seminar erzeugt einen virtuellen Handlungsraum, in dem die Akteure jederzeit “on-line” in das Simulationsgeschehen eingreifen können. Sie werden dort mit den für die gesellschaftliche Realität typischen komplexen und vernetzten Handlungssituationen konfrontiert und vertraut gemacht.

Konkret übernehmen die Seminarteilnehmer die Rollen verschiedener Prototypen von gesellschaftlichen Akteuren, nämlich Regierung, Opposition, Militär und Unternehmer, in einer von elf vorgegebenen Regionen. Auf der

Basis realer Daten für die wichtigsten ökonomischen und politischen Größen, die die Verhältnisse in unserer Welt auf dem aktuellen Stand abbilden, besteht die Aufgabe der Spieler darin, eine positive Entwicklung der eigenen Region zu bewirken.

Instrumente der Einflussnahme sind dabei u. a. die Bestimmung der gesamten Innenpolitik (Steuerfestlegung, Bildungs- und Wirtschaftsförderung, Festlegung gesellschaftspolitischer Richtlinien, etc.) durch den Regierenden, Kauf-, Investitions- und Personalentscheidungen durch den Unternehmer sowie militärische Aktionen (Truppenbewegungen bis hin zu Angriff auf andere Regionen oder innerregionaler Putsch) durch den Militärspieler. Unterstützt wird auch der Abschluss interregionaler Ver-

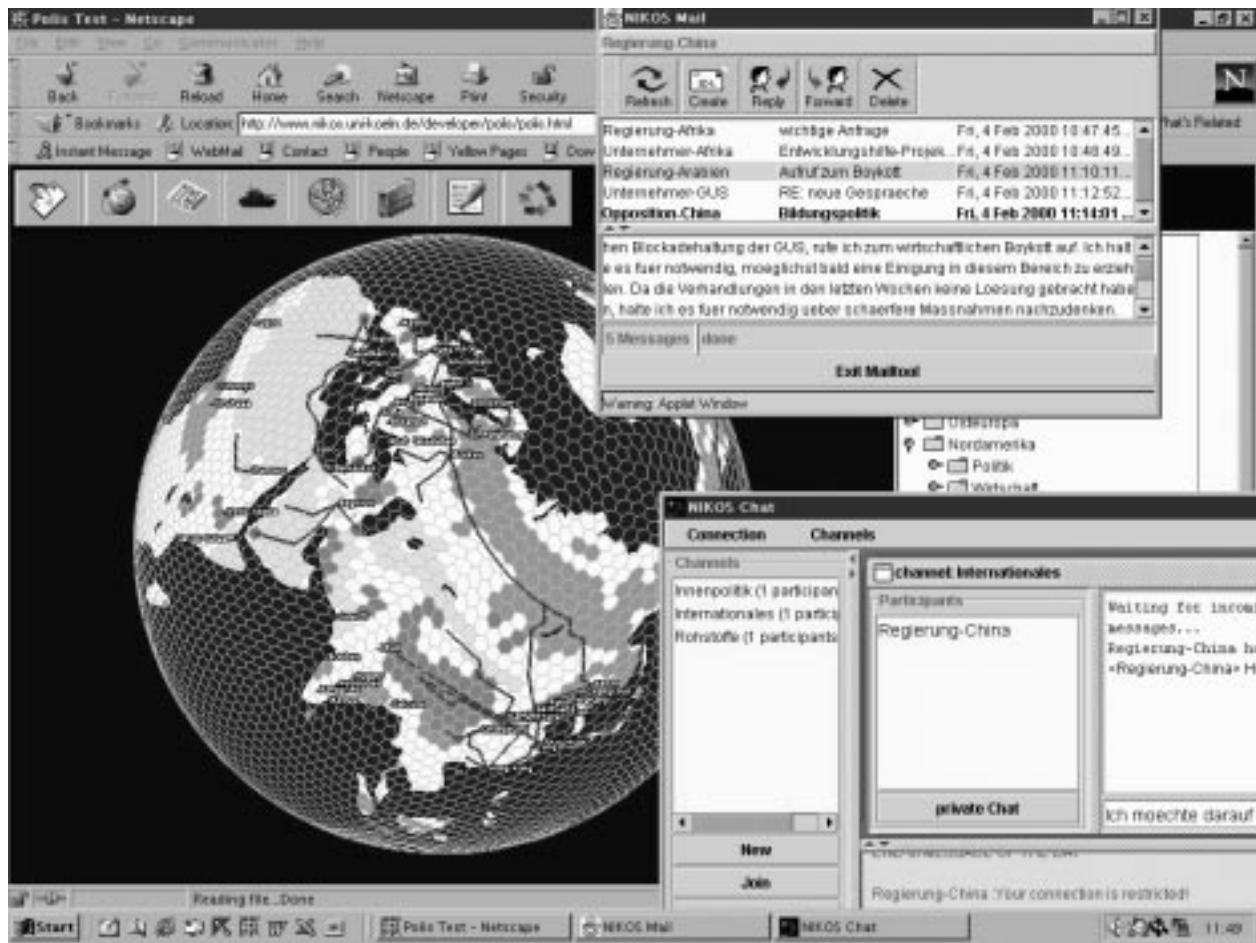


Abbildung 5.5: Screenshot einer Spielsituation für den Regierungsspieler der Region China.

träge.

Eine der zentralen Einsichten, die durch NIKOS vermittelt werden soll, ist die Notwendigkeit, nicht nur zwingende Entscheidungen zu treffen, sondern auch in wichtigen Fragen den Meinungs- und Informationsaustausch mit den anderen Akteuren zu suchen. Dafür wird eine Reihe von Komponenten wie Mail oder Chat zur Verfügung gestellt, die zur Gewährleistung größtmöglicher Plattformunabhängigkeit in Java realisiert sind. Darüber hinaus ermöglicht NIKOS den Erwerb von Schlüsselqualifikationen im Bereich des strategischen Denkens und Handelns auf spielerische Weise. Die Teilnehmer werden ermuntert, interaktiv den Verlauf der Lehrveranstaltungen mit zu gestalten.

NIKOS ist konzeptionell eingebettet in ein großflächiges Pilotprojekt der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln. Das Projekt wurde am ZAIK in Zusammenarbeit mit dem Seminar für Politikwissenschaft bei Herrn Prof. Leidhold realisiert. Er

wurde für das Plan- und Rollenspiels „Politik und internationale Sicherheit“ (POL&IS), das dem Projekt NIKOS zugrunde liegt, kürzlich mit dem Karl Carstens-Preis des Freundeskreis der Bundesakademie für Sicherheitspolitik e.V. ausgezeichnet.

Kontakt: nikos@zpr.uni-koeln.de

Kapitel 6

Modellierung und Simulation von Bausparkollektiven

Seit einigen Jahren besteht eine enge Zusammenarbeit zwischen den Landesbausparkassen (LBS) und unserer Arbeitsgruppe am ZAIK. Die Schwerpunkte der Kooperation bestehen sowohl in der Analyse großer Datenmengen, als auch in der Simulation komplexer sozioökonomischer Systeme. Die Ergebnisse werden von den Landesbausparkassen für ihre langfristige Geschäftsplanung verwendet.

Im Folgenden wird unser im Einsatz befindliches Simulationsmodell vorgestellt. Nach einem generellen Überblick werden die jüngsten Weiterentwicklungen des Modells beschrieben. In den beiden letzten Abschnitten werden zwei alternative Modellierungsansätze erläutert, die im Rahmen mehrerer Dissertationen entwickelt worden sind.

6.1 Ein deterministisches Modell für Bausparkollektive

6.1.1 Das Bausparen und unsere Tätigkeit – eine Übersicht

Der Ablauf eines Bausparvertrages gliedert sich für den individuellen Bausparer¹ im wesentlichen in zwei Phasen: Der Ablauf beginnt mit dem Abschluss eines Bausparvertrages über eine bestimmte Bausparsumme, die der Bausparer in sein Immobilienprojekt investieren will. In der Sparphase zahlt der Bausparer auf sein Konto ein und stellt seine Bauspareinlagen der Bausparkasse zur Verfügung. Mit der Zuteilung endet die Sparphase. Danach kann sich der Bausparer die Bausparsumme auszahlen lassen. Der Zeitpunkt der Zuteilung hängt davon ab, für

¹Zur Vereinfachung der Lesbarkeit unseres Artikels verwenden wir die männliche Form, wenn sowohl Frauen als auch Männer als Bausparer gemeint sind.

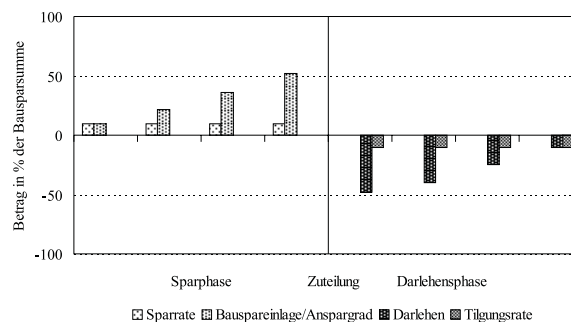


Abbildung 6.1: *Idealisierter Kontoverlauf mit konstanter Sparzahlung und konstanter Tilgung. Der Anspargrad ist der Anteil der Bauspareinlage an der Bausparsumme. Die Bewertungszahl (nicht abgebildet) ist ein abstraktes Maß für die bisher erbrachte Sparleistung.*

welchen Zeitraum der Bausparer welchen Betrag der Kasse zur Verfügung gestellt hat. Die Differenz zwischen der Bausparsumme und der Bauspareinlage wird dem Bausparer als Darlehen von der Kasse zur Verfügung gestellt. Mit der Auszahlung der Bausparsumme beginnt die Darlehensphase. In der Darlehensphase tilgt der Bausparer sein Darlehen mit einer vorgegebenen Mindesttilgungsrate (Abb. ??). Der Bausparer kann durch Aktionen wie vorzeitige Kündigung innerhalb der Sparphase, Erhöhung und Ermäßigung der Bausparsumme, Fortsetzung der Sparphase und Sondertilgung des Darlehens den Prozess von seiner Seite aus steuern.

Die Gesamtheit aller Bausparer wird als Bausparkollektiv bezeichnet. Die Funktion der Bausparkasse in dem Prozess besteht darin, dass sie innerhalb des Bausparkollektivs die Bauspareinlagen der einen Bausparer zusammen-

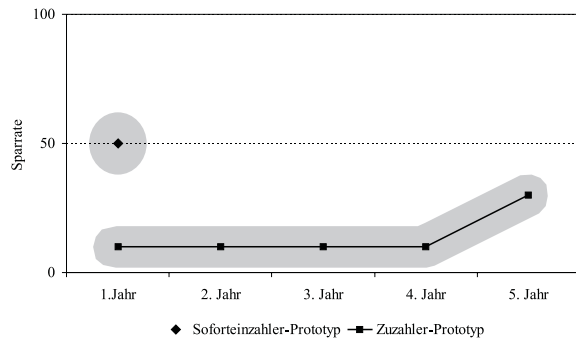


Abbildung 6.2: **Clusterung** mit zwei Clustern (graue Flächen) und dazugehörigen Prototypen. Der Prototyp *Sofortinzahler* repräsentiert das Verhalten, mit einer einmaligen Zahlung sofort den Mindestanspargrad zu erreichen. Der Prototyp *Zuzahler* repräsentiert das Verhalten, mit einer Erhöhung der Sparzahlung die Sparphase zu beenden.

fasst und als Darlehen an die anderen Bausparer verteilt. Für die Geschäftsplanung der Bausparkasse sind die Entwicklung der gesamten Bauspareinlagen, also die Überlagerung aller individuellen Einzahlungen, und die Entwicklung der zur Verfügung gestellten Darlehen, also die Überlagerung aller ausstehenden Darlehen, von zentraler Bedeutung. Unsere Arbeitsgruppe analysiert das Verhalten von Bausparkkollektiven in der Vergangenheit und erstellt aufgrund der gewonnenen Ergebnisse Prognosen für die zukünftige Entwicklung der Kollektive. Außerdem erstellen wir auf der Basis vorgegebener Eckdaten der Bausparkassen Szenarien für die zukünftige Entwicklung von Bausparkkollektiven. Dabei sind alle Problemebenen in die Arbeitsgruppe integriert: Die zugrunde liegenden statistischen Modelle und die darauf aufbauenden Algorithmen werden hier entwickelt. Anschließend werden die Algorithmen implementiert und auf die Kollektivdaten, welche die Bausparkassen uns bereitstellen, angepasst und angewendet. Auftretende Probleme mit den Daten und neue Anforderungen seitens der Bausparkasse fließen somit unmittelbar in die Weiterentwicklungen ein, welche auf den unterschiedlichen Problem instanzen getätigt werden.

6.1.2 Modellansatz

Die statistische Darstellung des Verhaltens eines Bausparkkollektives lässt sich in zwei Schritte unterteilen. Im

ersten Schritt wird ausschließlich das Sparverhalten betrachtet. Der einzelne Bausparer kann seine Sparzahlungen frei gestalten, was zur Folge hat, dass die Menge aller Sparverhaltensweisen des Bausparkkollektivs sehr inhomogen ist. Der statistische Zugang erfolgt über die Clusteranalyse: Zuerst werden diejenigen Verträge aus dem Bestand von Bausparverträgen herausgegriffen, von denen eine vollständige Sparphase vorliegt. In der Clusteranalyse werden die Sparraten, also die Sparzahlungen relativ zur Bausparsumme des jeweiligen Vertrages, der einzelnen Verträge miteinander verglichen. Die Verträge mit ähnlichem Sparverhalten werden zu einem Cluster zusammengefasst, dessen mittlere Sparrate, Prototyp genannt, ermittelt wird (Abb. ??). Die absoluten Sparzahlungen des Clusters ergeben sich aus dem Produkt der Sparraten des Prototypen mit der Bausparsumme des Clusters.

Anschließend werden die übrigen Verträge, bei denen die Sparzahlungen noch nicht beendet sind, denjenigen Prototypen zugeordnet, die im Sparverhalten den Verträgen am ähnlichsten sind. Außerdem wird jedem Prototypen die Bausparsumme aller derjenigen Verträge, die dem jeweiligen Prototypen zugeordnet wurden, zugewiesen. Der Prototyp repräsentiert sowohl das historische Sparverhalten der Gesamtheit der zugeordneten Verträge, als auch das zukünftige Sparverhalten der Gesamtheit (Abb. ??).

Im zweiten Schritt wird das weitere Verhalten, das innerhalb des Vertragsablaufs auf das Sparverhalten folgt, in die statistische Darstellung des Bausparkkollektivs miteinbezogen. Sowohl das Tilgungsverhalten als auch die anderen Aktionen, mit denen der Bausparer den Bausparprozess steuern kann, werden durch bestimmte Bausparaktionen beschrieben. Die Bausparaktionen spannen in ihrer

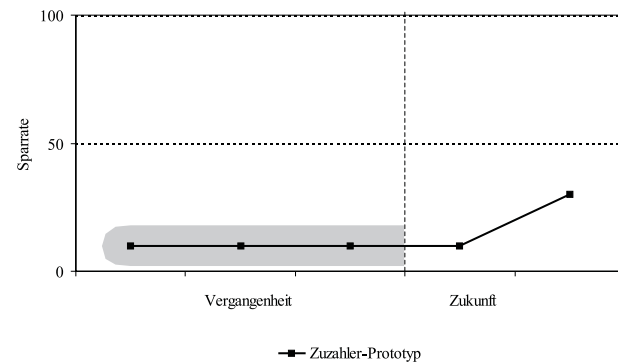


Abbildung 6.3: **Zuordnung** von Verträgen mit noch nicht beendeten Sparzahlungen (graue Fläche) zu dem *Zuzahler-Prototypen*.

kausalen Reihenfolge einen Entscheidungsbaum auf. Als Schicht wird die Kombination eines Prototypen mit einer vollständigen Folge von Bausparaktionen, die also bis zu einer Spitze des Entscheidungsbaums führt, bezeichnet. Jeder Prototyp spaltet sich über die Verzweigungen des Baumes in Schichten auf (Abb. ??).

Die Bausparsumme, die dem Prototyp zugewiesen wurde, verteilt sich ebenfalls entlang den Verzweigungen des Baumes auf die einzelnen Schichten. Anhand derjenigen Verträge, die dem Prototypen zugeordnet wurden, werden die Anteile bestimmt, mit der sich die Bausparsumme auf die verschiedenen Verzweigungen verteilt. Auf diese Weise gelangt man zu der Bausparsumme einer einzelnen Schicht. Das vollständige Verhalten eines Bausparkollektivs wird durch die Schichten und deren Bausparsummen abgebildet. Die Simulation eines einzelnen Bausparvertrages besteht darin, dass aus den primären Größen, Sparzahlungen und Bausparaktionen, die sekundären Größen wie Guthaben, Darlehen, Zinsen, Gebühren usw. in ihrem Zeitverlauf berechnet werden. Die Simulation eines Bausparkollektivs setzt sich aus den Simulationen einzelner Schichten zusammen, die unabhängig voneinander durchgeführt werden (Abb. ??). Das Simulationsmodell ist insofern deterministisch, als mit der Bestimmung der Schichten und deren Bausparsummen das zukünftige Verhalten des Bausparkollektivs eindeutig festgelegt ist.

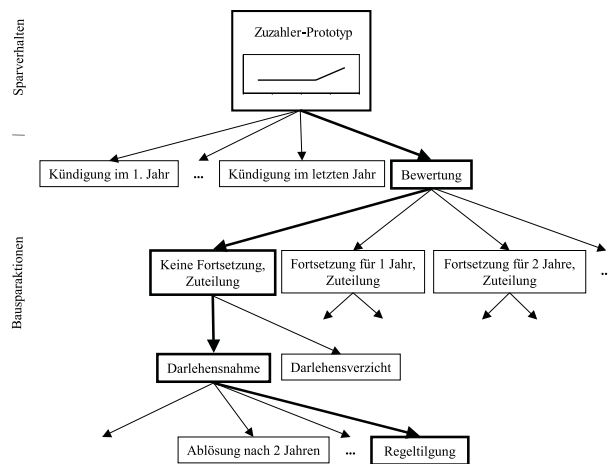


Abbildung 6.4: Aufspaltung des Zuzahler-Prototypen in **Schichten**. Eine Schicht ist als Beispiel durch Fettdruck hervorgehoben.

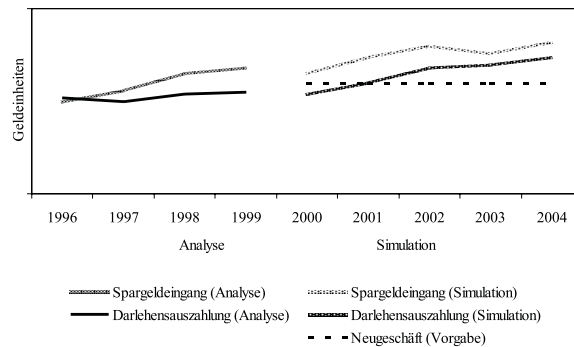


Abbildung 6.5: Idealisierte Darstellung der **Ergebnisse** einer Analyse und einer anschließenden Simulation unter Vorgabe eines konstanten Neugeschäfts.

6.2 Weiterentwicklung des Modells

Inzwischen hat sich der Modellansatz bei zahlreichen Simulationen für mehrere Landesbausparkassen bewährt, jedoch gab der praktische Einsatz des Modells auch Ansatzpunkte für Verbesserungen, hauptsächlich in puncto Komplexität und Steuerbarkeit. Die Notwendigkeit der Komplexitätsreduktion im Modell ergibt sich angesichts ausgereizter Rechnerressourcen aus dem sich verändernden Bauspargeschäft, bei dem in Zukunft weitere Tarifkonstruktionen und Handlungsmöglichkeiten für die Bausparer abzubilden sind, ebenso wie aus der komplizierten Handhabbarkeit des Modells. Daneben sollte die Durchführung von Simulationen effizienter gemacht werden, als auch der seitens der Bausparkassen nötige Aufwand gesenkt werden, der mit der Bereitstellung der Kollektivdaten verbunden ist.

6.2.1 Datenbank

Zur Ermittlung der Häufigkeiten, mit denen Bausparer bestimmte Aktionen (z. B. Kündigung, Fortsetzung, Darlehensverzicht etc.) durchführen, ist eine möglichst genaue Untersuchung der Einzelvertragsdaten im Zeitverlauf notwendig. Seit diesem Jahr wird ein relationales Datenbankmanagement-System von uns eingesetzt, so dass Abfragen wesentlich schneller in SQL formuliert werden können. Durch Schnittstellen zum Cluster- und Zuordnungsprogramm können nun zudem auch relativ

leicht prototypenspezifische Häufigkeiten ermittelt werden, da hierdurch in der Datenbank bekannt ist, zu welchem Prototyp ein bestimmter Vertrag zugeordnet wurde.

6.2.2 Erzeugung neuer Prototypen

Um die Anzahl der Prototypen und damit auch die der sich daraus ergebenden Schichten zu senken, wurden neue Prototypen erzeugt, wobei im Gegensatz zur ursprünglichen Clusterung nach Tarifen getrennt geclustert wurde. Diese Vorgehensweise erlaubt es, ein gutes Clusterergebnis auch mit deutlich weniger Clustern bzw. Prototypen zu erzielen, da das Sparverhalten innerhalb ein und desselben Tarifes deutlich homogener ist als über mehrere Tarife. So existieren z. B. in Tarifen mit 40% Mindestanspargrad keine 50%–Sofortinzahler und umgekehrt. Daneben wurden zusätzlich zu den Verträgen, die ihre Sparphase komplett beendet haben, auch Kündigungsverträge geclustert. Diese leisten einen teilweise erheblichen Beitrag zum Spargeldeingang bis zu ihrer Kündigung, und einige Tarife bestehen fast ausschließlich aus Kündigern. Ein weiterer Vorteil der Neuclusterung besteht in der höheren Aktualität und Vollständigkeit des Datenmaterials: Da die Daten der Jahre 1997 und 1998 hinzukamen, sind die Vertragsverläufe über 14 statt über nur 12 Jahre bekannt, so dass insbesondere Langsparere besser abgebildet werden konnten.

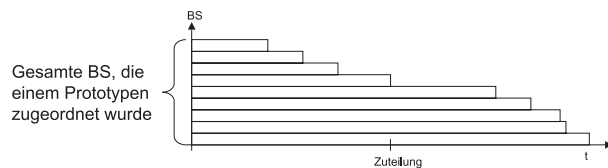


Abbildung 6.6: *Separate Berechnung einzelner Schichten im NBI.*

6.2.3 Dynamische Schichtenabspaltung und Steuerung

Das Simulationsprogramm NBI, welches die einzelnen Schichten von Vertragsanfang bis -ende durchrechnet, war ursprünglich zur Simulation einzelner Bausparverträge konzipiert. Eine Kollektivsimulation bestand aus der separaten Simulation vieler einzelner Schichten. Die große Zahl an möglichen Bausparaktionen, die einem Bausparer

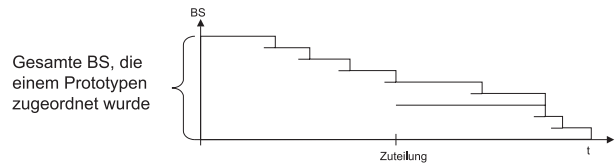


Abbildung 6.7: *Abspaltung der einzelnen Schichten im Simulationsverlauf.*

zur Verfügung steht, führt allerdings zu einem überproportionalen Anstieg der Schichten, in die die einem Prototyp zugeordnete Bausparsumme (in der Grafik Abb. ?? mit BS bezeichnet) zerlegt und anschließend separat durchgerechnet werden muss, so dass angesichts weitgehend ausgereizter CPU- und Speicherkapazitäten einer Berücksichtigung weiterer Wahlfreiheiten mit dem alten Ansatz enge Schranken gesetzt waren.

Zwar konnte das Problem des überproportionalen Anstiegs der Schichtanzahl mit steigenden Optionsmöglichkeiten nicht prinzipiell gelöst werden, allerdings ermöglicht ein geschickterer Algorithmus einen erheblichen Effizienzgewinn. Dieser beruht auf der Idee, dass eine weitere Verteilung der einem Prototypen zugeordneten Bausparsumme auf Schichten in der Simulation solange verzögert werden kann, bis das abweichende Sonderverhalten auch tatsächlich berücksichtigt werden muss (siehe Abb. ??). Die entsprechende Schicht wird erst zu diesem Zeitpunkt von der Bausparsumme des Prototypen abgespalten. Damit wird immer mit der minimal notwendigen Anzahl Schichten gerechnet. Hinzu kommt, dass Schichten, die kündigen, auf ihr Darlehen verzichten oder das Darlehen ablösen, sofort nach ihrer Abspaltung auch wieder verschwinden, da diese zuende berechnet sind.

Damit ist wieder für Kapazitäten gesorgt, die zur Abbildung zukünftiger Tarifkonstruktionen mit weiteren Handlungsmöglichkeiten für die Bausparer nötig sind.

Im Zuge der Simulationen für die Kassen stellte sich heraus, dass die Erstellung bestimmter Szenarien die simulationszeitabhängige Variation einiger Verhaltenshäufigkeiten erfordert, um auf diese Weise z. B. Trends leichter abbilden zu können. Im Rahmen der Umsetzung der dynamischen Schichtabspaltung konnte diese Anforderung durch eine zeitabhängige Skalierung der abzuspaltenden Bausparsummen relativ leicht umgesetzt werden, so dass der künftige Aufwand zur Erzeugung bestimmter Szenarien deutlich geringer ausfallen wird.

6.2.4 Zuordnung von „kurzen Kassen“

Die Anwendung des Simulationsmodells erfordert auch von den zu simulierenden Bausparkassen einen hohen Aufwand an Vorarbeit. Da die einzelnen Verträge eines Kollektivs den Prototypen zugeordnet werden müssen, werden Daten über das Verhalten jedes einzelnen Bausparvertrages im Zeitverlauf über einen möglichst langen Zeitraum benötigt. Diese Daten zu beschaffen stellt insbesondere für Kassen, die bislang noch keine Daten geliefert haben, einen erheblichen Aufwand dar, von dem sich ein großer Teil einsparen lässt, wenn lediglich Daten über einen kurzen Zeitraum bis hin zu den Konto-Endständen des vergangenen Jahres geliefert werden können. Da eine Clusterung auf der Basis solcher „kurzer“ Daten nicht sinnvoll ist, muss auf die aus den Daten einer anderen Kasse gewonnenen Prototypen zurückgegriffen werden. Die Zuordnung solcher „kurzer“ Daten zu den Prototypen ist allerdings aus dem Grund kritisch, da die Zuordnung der einzelnen Verträge durch einen Vergleich der Sparraten im Zeitverlauf mit denen der Prototypen erfolgen sollte, nämlich mit demselben Abstandsmaß, mit dem auch die Clusterung durchgeführt wurde. Da sich aber sowohl die Höhe als auch die zeitliche Verteilung der Sparraten im aktuellen Anspargrad und der Bewertungszahl eines Bausparvertrages niederschlagen, können diese Größen stellvertretend für die Sparraten zur Zuordnung verwendet werden. Hierzu werden die entsprechenden Werte für jeden Prototyp zu jedem Zeitpunkt berechnet und ein Abstandsmaß definiert, welches beide Größen so gewichtet, dass die Übereinstimmung mit der herkömmlichen Zuordnung am größten ist. Insgesamt zeigte sich bei Vergleichssimulationen mit denselben Daten die prinzipielle Eignung dieses Vorgehens.

6.3 SVD-Clusterung von Bauspar-Zeitreihen

Für die Prototypengenerierung haben wir bisher ein geometrisches Clusterverfahren, das K -means-Verfahren, verwendet, das diejenigen Verträge zu einem Cluster zusammenfasst, deren Sparverhalten im Zeitverlauf ähnlich sind. Als Ähnlichkeitsmaß haben wir den euklidischen Abstand gewählt. Der Prototyp eines Clusters wird durch den Clusterschwerpunkt definiert. Mit Hilfe dieses Verfahrens konnten wir Prototypen finden, die das Bausparkollektiv gut abbilden und auch aus der Realität bekannt

sind. Alternativ zu dem K -means-Clusterverfahren haben wir eine neue Methode zur Generierung der Prototypen untersucht, die auf der Singular Value Decomposition (SVD) basiert und kein explizites Abstandsmaß verwendet. Unter der Singular Value Decomposition wird die Zerlegung einer Matrix A in das Produkt zweier orthogonaler Matrizen U und V und einer Diagonalmatrix D verstanden: $A = UDV^T$. Als Ausgangsmatrix haben wir die Bausparmatrix A so definiert, dass die Zeilen die Verträge und die Spalten die Bausparmerkmale wie Sparrate (Verhältnis zwischen Spargeldeingang und Bausparsumme), Anspargrad (Verhältnis zwischen angespartem Guthaben und Bausparsumme) oder Bewertungszahl (Kennzahl für die Leistung des Bausparers für das Bausparkollektiv) im Zeitverlauf darstellen. Die Bausparmatrix kann um weitere Merkmale wie demographische Größen (Beruf oder Alter des Bausparers) erweitert werden. Auch hier haben wir als Datenbasis alle Bausparverträge betrachtet, deren Sparphasen uns vollständig bekannt sind. Bilden wir die Singular Value Decomposition der normierten Bausparmatrix und betrachten die zu den d größten Werten der Diagonalmatrix D gehörenden Spalten der linken orthogonalen Matrix U , dann erhalten wir aus den zu den p größten positiven (negativen) Spalteneinträgen gehörenden Verträgen pro Spalte $2d$ Cluster. Die Prototypen berechnen sich als Schwerpunkte der gefundenen Cluster. Zur Berechnung der Prototypen haben wir die relationale Bauspar-Datenbank eingesetzt, mit deren Hilfe wir die Cluster und die Prototypen leicht bestimmen können. Die mittels SVD-Clusterung gefundenen Prototypen weisen einen ähnlichen Sparverlauf wie die Prototypen der K -means-Clusterung auf und können auch in der Realität wiedergefunden werden. Somit haben wir hier ein zur K -means-Clusterung alternatives Clusterverfahren an der Hand, mit dessen Hilfe wir auch weitere Merkmale außer der Sparrate zur Bestimmung der Prototypen hinzuziehen können, ohne einen expliziten Abstand zu berechnen.

6.4 Stochastische Modellierung mit Hidden-Markov-Modellen

Ein alternatives Vorgehen zur deterministischen Modellierung von Bausparkollektiven, wie es in den vorherigen Abschnitten beschrieben wurde, besteht in der stochastischen Beschreibung des möglichen Verhaltens von Bausparern. Im Rahmen zweier Dissertationen wurde hierzu

ein Ansatz entwickelt und untersucht, der auf sogenannten Hidden-Markov-Modellen (HMM) basiert. Zu diesem Zweck wurde eine umfangreiche, universelle HMM-Bibliothek zur Modellierung heterogener, nicht stationärer Daten implementiert, die auch in Bereichen außerhalb des Bausparwesens verwendet werden kann. Hidden-Markov-Modelle werden seit einigen Jahren sehr erfolgreich in verschiedensten Anwendungsgebieten wie z. B. in der Bioinformatik, der Mustererkennung, der automatischen Spracherkennung und der Zeitreihenanalyse eingesetzt.

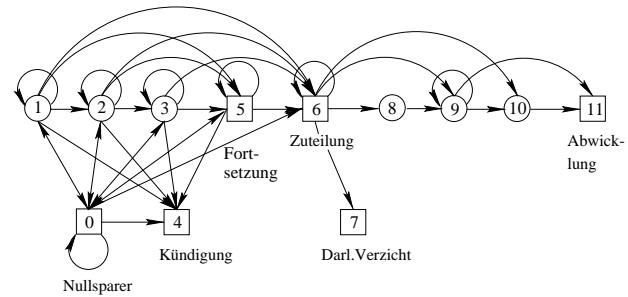


Abbildung 6.9: Graph zur HMM-Modellierung eines Bausparvertrages.

6.4.1 Hidden-Markov-Modelle

Ein HMM beschreibt einen stochastischen Prozess, der sich aus zwei gekoppelten Mechanismen zusammensetzt: Eine „versteckte“ Markov-Kette mit einer endlichen Anzahl von Zuständen wird in diskreten Zeitschritten durchlaufen und generiert dabei in jedem Zustand ein Ausgabesymbol gemäß einer von dem jeweiligen Zustand abhängenden Wahrscheinlichkeitsdichte. Für einen Beobachter ist nur die so entstehende Sequenz von Ausgabesymbolen sichtbar, während die darunterliegende Folge von Zuständen verborgen bleibt.

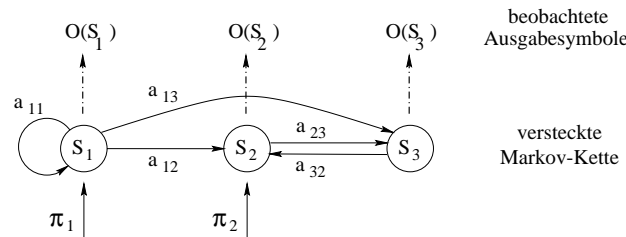


Abbildung 6.8: Beispiel für ein einfaches HMM mit drei Zuständen.

Abbildung ?? zeigt ein Beispiel für ein einfaches HMM mit den relevanten Modellparametern. Die Knoten des Graphen, symbolisiert durch Kreise, stehen für die möglichen Zustände, die das System einnehmen kann, und die gerichteten Kanten entsprechen jeweils den Übergängen zwischen zwei Zuständen, wobei jede Kante mit der jeweiligen Übergangswahrscheinlichkeit (a) gewichtet ist. Die zusätzlichen Eingangskanten unter den Knoten enthalten die Wahrscheinlichkeit (π), dass der Prozess in dem entsprechenden Zustand startet, und $O(S)$ kennzeichnet die Wahrscheinlichkeitsdichte, mit welcher im Zustand S die Ausgabesymbole generiert werden.

6.4.2 Modellansatz

Was hat jetzt ein solch abstraktes Modell mit dem Verhalten von Bausparern zu tun? Die Idee besteht darin, dass auch ein Bausparer während der Vertragslaufzeit verschiedene Phasen durchläuft, die zum Teil von seinen persönlichen Gegebenheiten und Intentionen abhängen und zum Teil durch äußere Faktoren festgelegt werden, die mit der Bausparkasse und dem typischen Ablauf eines Bausparvertrages zusammenhängen. Die persönlichen Umstände sind selbstverständlich schwer fassbar und in der Regel auch in den zur Verfügung stehenden Daten nicht enthalten. Dies steht im Einklang mit der Unbeobachtbarkeit der Zustände eines Hidden-Markov-Modells. Die Ausgabesymbole des Modells dagegen sind beobachtbar und entsprechen den Aktionen des Bausparers, wie z. B. Spareinzahlungen, Tilgungszahlungen, Kündigung oder Darlehensverzicht. Die zeitliche Abfolge dieser Aktionen wird im Modellansatz als eine Sequenz von Ausgabesymbolen aufgefasst, und das gesamte Bausparkollektiv besteht in dieser Modellierung folglich aus einer entsprechend großen Anzahl einzelner Sequenzen.

In Abbildung ?? ist die Topologie des in der Bausparsimulation eingesetzten Gesamtmodells dargestellt. Zustände, die mit einem Quadrat gekennzeichnet sind, spielen eine besondere Rolle, da sie lediglich ein fest definiertes, den jeweiligen Zustand kennzeichnendes Symbol ausgeben können. Hierdurch ist es möglich, Zustände fest mit bestimmten Aktionen zu verknüpfen. Außerdem wird durch die Einschränkung der Übergänge der prinzipielle Ablauf eines Bausparvertrages vorgegeben.

Ein Grund für den verbreiteten Einsatz von Hidden-Markov-Modellen ist darin zu sehen, dass mit dem Baum-Welch-Algorithmus ein effektives Verfahren vorhanden ist, das ein Training der Modelle und damit eine Anpas-

sung an vorhandene Daten ermöglicht. Die vom Algorithmus ermittelten Modellparameter stellen ein lokales Optimum im Parameterraum dar. Sie werden also durch das Training so ermittelt, dass die vorgegebenen Daten möglichst gut vom Modell repräsentiert werden.

6.4.3 Training und Erweiterung der Hidden-Markov-Modelle

Das Training der Hidden-Markov-Modelle geschieht mit Bausparverträgen, deren Spar- oder Darlehensphase vollständig abgeschlossen ist. Dabei werden mehrere, zunächst recht unspezifische Modelle vorgegeben, die sich im Laufe des Trainings auf eine Teilmenge der Verträge spezialisieren können. Dies stellt eine Analogie zur in den vorhergehenden Abschnitten beschriebenen Prototypenbildung durch Clusterung dar, wobei diese Teilmengen den Clustern entsprechen und die Parameter der trainierten Hidden-Markov-Modelle den Prototypen. Im deterministischen Bauspar-Modell, wie in den vorhergehenden Abschnitten beschrieben, ergeben sich die Prototypen durch eine Mittelung über die Verträge im Cluster. Im Fall der Hidden-Markov-Modelle werden die statistischen Eigenschaften der Sequenzen durch die Modellparameter abgebildet, die entstehenden Prototypen sind also stochastischer Natur. Um diese stochastischen Prototypen zu gewinnen, wird das Training der Modelle durch den Baum-Welch-Algorithmus mit zwei bekannten Clusterverfahren, dem K-Means-Verfahren und dem EM-Algorithmus zur Clusterung mit Mischmodellen, kombiniert.

Um Hidden-Markov-Modelle erfolgreich zur Simulation von Bauspardaten einzusetzen, wurden die Standardmodelle in einigen Punkten erweitert. Dies umfasst unter anderem die Integration von deterministischen Nebenbedingungen, die aus gesetzlichen und tariflichen Vorgaben resultieren. Desweiteren ist es notwendig, Sequenzen mit einer Gewichtung zu versehen, da in der Realität Verträge immer mit einer Bausparsumme verbunden sind. Außerdem muss durch die Wahl einer geeigneten Wahrscheinlichkeitsdichte für die Ausgaben sichergestellt werden, dass keine negativen Spar- bzw. Tilgungszahlungen vom Modell erzeugt werden, da dies in der Realität nicht vorkommt.

6.4.4 Simulationen

Sind die optimalen Modellparameter gefunden, so ist es über einen Monto-Carlo-Ansatz sehr einfach möglich,

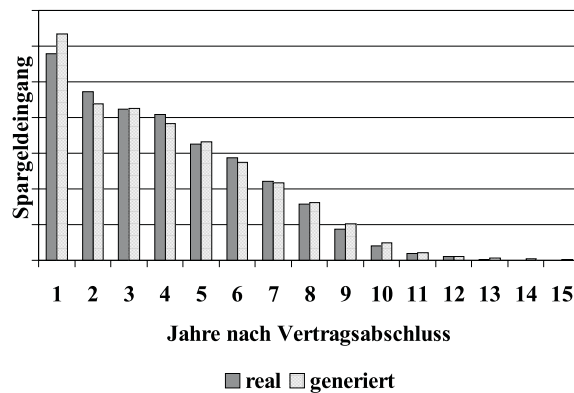


Abbildung 6.10: Verteilung der Spargeldeingänge bei realen Trainingsdaten und bei künstlichen, mit einem HMM generierten Sequenzen.

neue, künstliche Sequenzen zu generieren oder auch bestehende Teilsequenzen zu vervollständigen. Hierzu werden unter Verwendung eines Zufallszahlengenerators basierend auf den Modellparametern neue Markov-Ketten und dazugehörige Ausgabesymbole generiert. Die Anteile der verschiedenen Modelle an den künstlichen Sequenzen ergeben sich unmittelbar aus den jeweiligen Volumina der Modelle in der Clusterung.

Zur Verlängerung vorhandener Teilsequenzen muss zunächst bestimmt werden, welchem der zur Auswahl stehenden Hidden-Markov-Modelle die Sequenz entspringt. Da dies in der Regel nicht eindeutig möglich ist, wird hierfür das Modell mit der größten Wahrscheinlichkeit zur Erzeugung dieser Sequenz gewählt, wobei es über einen Bayes-Ansatz möglich ist, a priori Wahrscheinlichkeiten der verschiedenen Modelle zu berücksichtigen. Dies hat Relevanz bei einer evtl. eingreifenden Steuerung von Simulationen zur Durchführung von Szenarien und bei der Verlängerung sehr kurzer Teilsequenzen, bei denen die Zuordnung zu einem bestimmten Modell zwangsläufig mit einer gewissen Unsicherheit verbunden ist. Ist das Generatormodell bestimmt, so wird im nächsten Schritt der wahrscheinlichste Zustand dieses Modells zum Zeitpunkt des Sequenzendes berechnet, um hiervon ausgehend neue Symbole zu generieren.

6.4.5 Ergebnisse

Diverse Untersuchungen mit dem beschriebenen Modellansatz haben gezeigt, dass Hidden-Markov-Modelle sehr gut geeignet sind, die stochastischen Charakteristika der

Bauspar-Zeitreihen zu modellieren. Beispielhaft zeigt Abbildung ?? einen Vergleich der Spargeldaufkommen in den verschiedenen Jahren nach Vertragsabschluss von realen Trainingsdaten und einer künstlich generierten Menge von Sequenzen.

Die beschriebenen Erweiterungen führen zu einer deutlich verbesserten Abbildung der relevanten Größen gegenüber dem Standard-HMM. Erste Vergleiche eines generierten Gesamtkollektivs mit den entsprechenden realen Daten sind vielversprechend und lassen einen erfolgreichen Einsatz des Modells zur Durchführung von Simulationen in der Praxis erwarten.

6.5 Ausblick

Die Weiterentwicklung des Simulationsmodells hängt stark von den Zielsetzungen der LBS ab. Als wichtiger Punkt ist sicherlich die Implementierung eines flexibilisierten Bewertungs- und Zuteilungsverfahrens zu nennen. Ein solches Verfahren wird für das Simulationsmodell neben eher technischen Fragen der konkreten Realisierung insbesondere auch grundsätzliche Probleme der Bestimmung bestimmter Parameter aufwerfen. Hier stellt sich die Frage, wie bei neu einzuführenden Handlungsmöglichkeiten für die Bausparer die entsprechenden Häufigkeiten abgeleitet werden können. Interessant wären auch tiefergehende Untersuchungen im Zusammenhang mit der dynamischen Steuerung der Simulationen und hier insbesondere der Einfluss äußerer Faktoren auf das Kollektivverhalten. Mit der beschriebenen dynamischen Schichtenabspaltung wurde die Voraussetzung einer effektiven Steuerung geschaffen.

Im Rahmen der stochastischen Modellierung mit Hidden-Markov-Modellen ergeben sich sowohl aus theoretischer als auch aus praxisorientierter Sicht eine Reihe weiterer interessanter Fragestellungen und Aufgaben, die den Gegenstand einer wissenschaftlichen Arbeit bilden können. Im theoretischen Bereich ist sicherlich die Entwicklung von Verfahren zur Überwindung lokaler Maxima beim Training der Modelle und bei der Clusterung ein lohnenswertes Projekt. Einer Anwendung dieses Modells in der Praxis müsste auf jeden Fall die Kopplung mit dem bestehenden Rechenkern NBI zur Berechnung verschiedener sekundärer Größen wie Zinsen und Gebühren und zur Durchführung von Szenariorechnungen vorausgehen.

Kontakt: `bauspar@zpr.uni-koeln.de`

Anhang A

Computer und Netzwerke

Die Technik schreitet, wie man weiss, mit großen Schritten voran. Ein Narr, der sich dem aktuellen Trend unterwirft. Trotzdem kommt auch unsere Arbeitsgruppe nicht umhin, von Zeit zu Zeit einen Blick auf die Infrastruktur zu werfen und abzuwägen, an welcher Stelle es etwas hakt. Aus diesem Grunde stellen wir die aktuelle technische Konfiguration unseres Hauses vor.

A.1 Zentrale Server

Als zentrale Server kommen zur Zeit zwei Maschinen von SUN Microsystems zum Einsatz. Der mit Abstand größte Teil der über NFS zugreifbaren Festplatten hängt an einer Enterprise 4000 mit acht CPUs, die auch zum interaktiven Arbeiten verwendet wird. Für rechenintensive Anwendungen steht eine E450 mit vier schnellen CPUs zur Verfügung.

Um die Speicherung und Verwaltung der großen Datenmengen, wie sie z. B. innerhalb unserer Kooperation mit den Landesbausparkassen anfallen, bewältigen zu können, ist ein hierarchisches Speichersystem installiert. Das System besteht aus einer SUN Enterprise 3500 mit zwei CPUs und einem DLT-Bandroboter mit einer Speicherkapazität von 5 TeraByte. Als Software wird das modulare hierarchische Dateisystem EPOCH eingesetzt. Dieses Dateisystem erlaubt die tägliche Sicherung der Daten, so dass sich einfach ein gewünschter alter Datenzustand abrufen lässt (Backup), sowie die Datenmigration. Migration bezeichnet das Verfahren, die Daten geeignet zwischen Platten und Magnetbändern zu verteilen, so dass man auf die gesamten 5 TeraByte wie auf ein normales Dateisystem zugreifen kann. Häufig benutzte Daten liegen dafür auf Platte vor, seltener benutzte werden automatisch auf DLT-Tapes ausgelagert und bei Bedarf für den Benutzer transparent zurückgerufen.

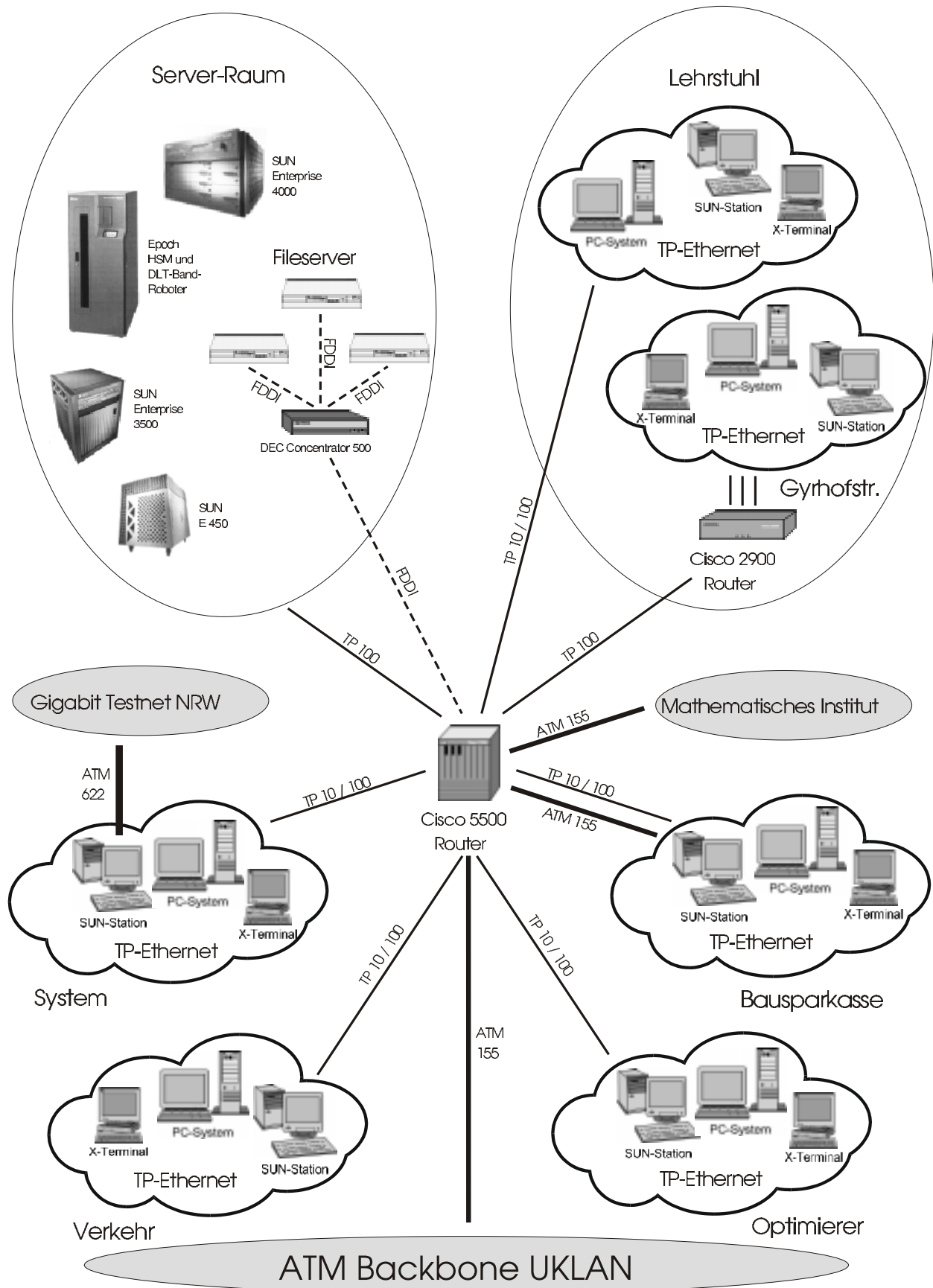
A.2 Arbeitsplatzrechner

Allen unseren Mitarbeitern, Habilitanden, Doktoranden und Diplomanden steht ein eigener Arbeitsplatzrechner zur Verfügung. Diese setzen sich aus X-Terminals, SUN- und IBM-Workstations sowie PCs unter GNU/Linux und Windows NT zusammen. Alle Rechner greifen über SAMBA auf ein gemeinsames Dateisystem und die Drucker zu.

A.3 Vernetzung

Die am ZAIK eingesetzten Rechner sind über ein heterogenes Netzwerk miteinander verbunden. Diese Heterogenität bezieht sich dabei sowohl auf Verkabelungstechnik als auch auf die eingesetzten Protokolle. Man findet Twisted-Pair-Kabel der Kategorie 5 (UTP5), das sowohl den Einsatz von Ethernet (10 MBit/s) und Fast-Ethernet (100 MBit/s) als auch von 155 MBit/s ATM-Verbindungen erlaubt, und Glasfaserkabel bzw. Lichtwellenleiter, worüber ebenfalls 155 bzw. 622 MBit/s ATM-Verbindungen möglich sind.

Darüber hinaus wird hausintern zur Zeit auch noch FDDI (Fiber Distributed Data Interface) verwendet, das, wie der Name schon sagt, ebenfalls Glasfaser nutzt. Diese Technik wird aber nach und nach durch Fast-Ethernet bzw. ATM abgelöst. Die Anbindung an das Rechenzentrum erfolgt über zwei 155 MBit/s ATM-Leitungen, die in Zukunft aber vielleicht durch Gigabit-Ethernet ersetzt werden.



Anhang B

Gäste

Forschungsaufenthalte

Prof. Dr. V. Izhutkin



Prof. Dr. Victor S. Izhutkin besuchte im Zeitraum Oktober-November 2000 das ZAIK im Rahmen eines Deutschland-Stipendiums.

Herr Izhutkin ist ein Vertreter der kontinuierlichen Optimierung. Mittels seiner vielen wissenschaftlichen Kontakte, insbesondere nach Deutschland, hat sich Herr Izhutkin hier und weltweit einen angesehenen wissenschaft-

lichen Namen erwerben können.

Er ist Inhaber eines Lehrstuhl an der Mari State University in der zwischen Wolga und Ural gelegenen Stadt Yoshkar-Ola. Vor seinem wissenschaftlichen Hintergrund und angesichts der diskret-mathematischen Problemkreise des ZAIK gab es für ihn die Möglichkeit eines vertieften Einblickes in andere Fragen und Methoden, insbesondere in diskrete Algorithmen. In seinem Oberseminar-Vortrag griff Herr Izhutkin gleichsam die stark rechnerorientierte Arbeit des ZAIK auf, indem er im Hinblick auf die Lehre das computerunterstützte Lernprogramm erläuterte, welches an seinem Lehrstuhl für die kontinuierliche Optimierung entwickelt worden ist und von den Studierenden eingesetzt wird.

Die in diesen Kölner Tagen begonnenen fachlichen Korrespondenzen und persönlichen Freundschaften mit ihm werden gewiss wieder aufgefrischt und vertieft werden.

Y. Shokin



Prof. Dr. Yuri Shokin besuchte im November 2000 das ZAIK. Der Kontakt ergab sich durch die langjährige Zusammenarbeit mit G.-W. Weber und insbesondere durch die Bereitschaft von Herrn Shokin, dem Hauptherausgeber des Journal of Computational Technologies, die Proceedings zum Workshop „Discrete Optimization — Structure and Stability

of Dynamical Systems“ in einem Spezialband seiner Zeitschrift erscheinen zu lassen. So werden referierte Arbeiten von N. Eissfeldt, Ch. Hagemeier, W. Hochstättler, G. Leitmann, St. Pickl, R. Schrader, G.-W. Weber u.a. hierin veröffentlicht.

Herr Shokin ist Leiter des Institute of Computational Technologies, Novosibirsk (Russland), welches etwa einem deutschen Max-Planck-Institut entspricht. Besuche von Gorbatschow und Putin zeigen die anerkannt professionelle Institutsführung an. Er selbst ist einer der führenden Numeriker in Sibirien, Mitglied der Akademie der Wissenschaften und ein weltweit anerkannter und geschätzter Fachmann auf dem Gebiet der Strömungsmechanik. Insbesondere hat er eine Pionierarbeit bei den Finite-Differenzen-Verfahren geleistet.

Angesichts der schwierigen Situation Russlands und insbesondere Sibiriens ist die Ermutung von Öffnungs- und Modernisierungsversuchen, wie sie Yuri Shokin unternimmt, überaus wichtig. Solch eine Ermutung bleibt keine Einbahnstrasse, wie der ZAIK-Kontakt zu ihm beweist.

P. Erdős

Dr. Peter L. Erdős kommt aus der traditionsreichen Schule der ungarischen Kombinatorik, die von Paul Erdős (mit dem Peter jedoch weder verwandt noch verschwägert ist) begründet wurde. Er promovierte 1982 unter G.H.O Katoana an der Eötvös-Universität in Budapest und hat seither verschiedene Stellen in Ungarn und längere Zeit in den Niederlanden bekleidet. Deutschland kennt er als Stipendiat der Alexander von Humboldt-Stiftung an der Universität Bonn. Zur Zeit leitet er als Senior Research Associate

eine Forschergruppe am Mathematischen Institut der Ungarischen Akademie der Wissenschaften in Budapest.

Peter Erdős Forschungsinteresse umfasst die gesamte Breite der Kombinatorik. In der Theorie betrifft dies insbesondere die Kombinatorik von Mengensystemen, die gewisse extremale Eigenschaften aufweisen (sollten), und geordnete Mengen. Bei den Anwendungen steht die Biomathematik im Vordergrund und dabei speziell phylogenetische Bäume und die Spektraltheorie evolutionärer Bäume.

Gastvorträge

R. KANNAN: *Linear Algebra and Web Search*, 28.6.1999

H.-J. PRÖMEL: *Approximationsalgorithmen für das Steiner-Problem in Graphen*, 5.11.1999

J. PENG: *New Complexity Analysis for semidefinite Optimization*, 9.11.1999

J. NESETRIL: *Nowhere zero flows, antiflows and 7776-cloring of planar graphs*, 12.11.1999

A. ZIEN: *Ein iteratives LP-basiertes Verfahren zur Parameter-Kalibrierung und seine Anwendung auf Protein Strukturerkennung durch Threading*, 4.7.2000

G.-W. WEBER: *Verallgemeinerte semi-infinite Optimierung*, 7.7.2000

S. FUJISHIGE: *Polybasic Polyhedra*, 12.9.2000

P. ERDŐS: *The Word Poset*, 5.10.2000

V. ISHUTKIN: *Verfahren der reduzierten Richtungen für nichtlineare Optimierungsprobleme*, 10.10.2000

T. HORVATH: *A combinatorial model for the protein secondary structure prediction problem*, 17.10.2000

G. LEITMANN: *Some remarks about a method of direct optimization and its application to a class of Nash Equilibria*, 26.10.2000

V. ISHUTKIN: *Linear Correction of Smooth Optimization Problem Solution with Small Data Perturbations on the Basis of Reduced Direction*, 27.10.2000

B. FUCHS: *Untersuchung der oberen und unteren Schranke der Approximationsgüte des 3-Max-Cut-Problems*, 31.10.2000

P. ERDŐS: *Cuts in Graphs*, 10.11.2000

Y. SHOKIN: *Discrete Optimization*, 1.12.2000

Anhang C

Veranstaltungen

Festkolloquium in memoriam Klaus Wagner

Am 30. Juni 2000 fand an der Universität zu Köln ein Festkolloquium in memoriam Klaus Wagner statt.

Am 6. Februar 2000 verstarb Herr Professor Dr.Dr.h.c. Klaus Wagner, der als Pionier der deutschen Graphentheorie gilt. Er hätte sein 90stes Lebensjahr am 31. März 2000 vollendet. Klaus Wagner promovierte 1934 bei K. Doerge an der Universität Köln. Nach der Habilitation 1949 und der Ernennung zum Professor 1956 war er bis 1970 an der Universität Köln als wissenschaftlicher Rat und Professor tätig. Er wurde 1970 an die Universität Duisburg berufen und wirkte dort bis zu seiner Emeritierung im Jahre 1976. An der Universität Köln wurde er 1971 zum Honorarprofessor ernannt und im Jahr 1997 zum Ehrendoktor an der Universität Duisburg.

In dem Festkolloquium, das von der Arbeitsgruppe Faigle/Schrader und Dr. B. Randerath (Institut für Informatik) organisiert wurde, ging zunächst Prof. Dr. R. Halin (Universität Hamburg) in seinem Vortrag auf den *Mensch und Mathematiker* Klaus Wagner ein. Prof. Dr. H. Sachs (TU Ilmenau) referierte über ein kombinatorisch-geometrisches Problem von Deuber und stellte in seinem Vortrag viele Bezüge zu den Ergebnissen von Wagner dar. In dem Abschlussvortrag sprach Prof. Dr. C. Thomassen (TU of Denmark) über *Embedding and Coloring Graphs on Surfaces* und präsentierte viele offene Probleme dieses Gebietes, daß von Klaus Wagner mit begründet wurde.

An der Veranstaltung nahmen 50 Wissenschaftler aus dem In- und Ausland teil, die bei der Abendveranstaltung mit dem Sohn von Klaus Wagner viele Erinnerungen austauschten.

Workshop Discrete Optimization — Structure and Stability of Dynamical Systems

In dem Seminarraum des ZAIK fand am 26.10.2000 der Workshop *Discrete Optimization - Structure and Stability of Dynamical Systems* statt. Im Rahmen der Vorträge wurden die Arbeitsschwerpunkte der einzelnen Arbeitsgruppen Küpper, Seydel und Faigle/Schrader sowie von Herrn PD Dr. Gerhard-W. Weber (Lehrstuhlvertretung im WS 00/01 am Mathematischen Institut) vorgestellt und über aktuelle Forschungsergebnisse berichtet. Bereichert wurde die Veranstaltung, die von Dr. Stefan Pickl organisiert wurde, durch die zwei Hauptvorträge von Prof. Dr. George Leitmann (University of Berkeley) sowie von Prof. Dr. Werner Krabs (TU Darmstadt), die zu den führenden Experten auf den Gebieten *Optimale Kontrolle und Steuerung zeit-diskreter Prozesse und Spiele* zählen. Bei dem gemeinsamen Abendessen konnte die Zeit genutzt werden, Herrn Leitmann zu seinem zurückliegenden 75. und Herrn Weber zu seinem vergangenen 40. Geburtstag zu gratulieren.

Die Beiträge der Tagung werden im Journal of Computational Technologies veröffentlicht werden.

Lehrveranstaltungen

SS 99

Informatik I

R. Schrader mit A. Noltemeier und V. Stausberg

Entscheidungsunterstützungssysteme und Datenbanken

E. Dahlhaus

Seminar über Submodulare Funktionen

W. Hochstättler

WS 99/00

Mathematische Programmierung I

U. Faigle mit M. Kloock

Informatik II

R. Schrader mit T. Epping

Seminar über Probabilistische Methoden in der diskreten Optimierung

U. Faigle, M. Jünger, R. Schrader, E. Speckenmeyer

SS 00

Mathematische Programmierung II

U. Faigle

Diskrete Mathematik

U. Faigle mit M. Hayer

Algorithmen zu diskreten Strukturen

R. Schrader mit M. Hayer

Künstliche Intelligenz

R. Schrader

Programmierpraktikum

R. Schrader mit Ch. Hagemeyer und D. Rübiger

Seminar über Algorithmen der Bioinformatik

U. Faigle, W. Hochstättler, R. Schrader mit A. Schönhuth und A. Schliep

WS 00/01

Mathematische Grundlagen des Operations Research

U. Faigle mit M. Peter

Diskrete Strukturen und Algorithmen der Bioinformatik

S. Pickl

Seminar über Semidefinite Programmierung

U. Faigle

Seminar über Kombinatorik und Diskrete Mathematik

U. Faigle und W. Hochstättler

Anhang D

Arbeitsergebnisse und Konferenzteilnahmen

Veröffentlichungen

MASSIMO BERNASCHI, FILIPPO CASTIGLIONE, SAURO SUCCI: Large-scale Cellular Automata simulations of the Immune System response, *Physical Review E* **2** (2000), S. 1851–1854

MASSIMO BERNASCHI, FILIPPO CASTIGLIONE, SAURO SUCCI: A high performance simulator of the Immune Response, *Future Generation Computer System* **3** (1999), S. 333–342

MASSIMO BERNASCHI, FILIPPO CASTIGLIONE, P. SEIDEN, SAURO SUCCI: Learning Cascade in the immune system dynamics: a numerical simulation, *Int. Conf. on Computational Physics (CCP98), Grenada Nov 1998* (1998), S. 13

ULRICH BLASUM, WINFRIED HOCHSTÄTTLER, PETER OERTEL: Steiner-Diagrams and k-Star-Hubs, *submitted to: Discrete Applied Mathematics* (2000)

ROLF BÖNING, CHRISTIAN GAWRON, STEPHAN HASSELBERG, STEPHAN ROSSWOG, PETER WAGNER: Computational Aspects in Traffic Simulation Problems, *to appear: Future Generation Computer System* (1999), S. 9

FILIPPO CASTIGLIONE: Diffusion and Aggregation in an Agent Based Model of Stock Market Fluctuations, *International Journal of Modern Physics C* **5** (2000)

FILIPPO CASTIGLIONE: Forecasting price increments using an artificial Neural Network, *Advances in Complex Systems* (2000), S. 1–12

FILIPPO CASTIGLIONE, RAS B. PANDEY, DIETRICH STAUFFER: Effect of Trading Momentum and Price Resistance on Stock Market Dynamics: A Glauber Monte Carlo Simulation, *Physica A (in press)* (2000)

FILIPPO CASTIGLIONE, FRANCESCO CASTIGLIONE: Estimating the Keratoconus Index from ultrasound images of the human cornea, *submitted to: IEEE Transactions on Medical Imaging* (1999), S. 3

NILS EISSFELDT, MARCUS METZLER, PETER WAGNER: Dynamic simulation of traffic and its environmental impacts, *Proceedings of 9th IFAC Symposium Control in Transportation Systems 2000* (2000), S. 6

ANDREAS ERDMANN, ANDREAS NOLTE, ANJA NOLTEMEIER, RAINER SCHRADER: Modeling and Solving the Airline Schedule Generation Problem, *submitted to: Math. in Industrial Systems* (1999)

ULRICH FAIGLE, MARCEL HUNTING, WALTER KERN: A Lagrangian relaxation approach to the edge-weighted clique problem, *to appear: European Journal of Operational Research* (1999)

SANDOR P. FEKETE, JOSEPH S. B. MITCHELL, KARIN WEINBRECHT: On the continuous Weber and k-means problems, *To appear in: Proceedings of the 16th Annual ACM Symposium on Computational Geometry, SoCG'00* (1999), S. 18

MATTHIAS HAYER, WINFRIED HOCHSTÄTTLER: Test Sets for Vertex Cover Problems (Extended Abstract),

Proceedings 6th Twente Workshop on Graphs and Combinatorial Optimization (1999)

WINFRIED HOCHSTÄTTLER, CHRISTOPHER MUES, PETER OERTEL: Algorithmen für Speditionsroutingprobleme mit Umlademöglichkeit, *to appear: Tagungsband der 6. Fachtagung Logistik in Magdeburg* (2000)

WINFRIED HOCHSTÄTTLER, JAROSLAV NESETRIL: A Note on MaxFlow-MinCut and Homomorphic Equivalence in Matroids, *to appear: Journal of Algebraic Combinatorics* (1999)

WERNER KRABS, STEFAN PICKL, JÜRGEN SCHEFFRAN: Optimization of an n-person game under linear side conditions, *Optimization, Dynamics and Economic Analysis – Essays in Honor of Gustav Feichtinger* (2000)

ERIC KROPAT, STEFAN PICKL, ANDREAS RÖSSLER, GERHARD-WILHELM WEBER: A New Algorithm From Semi-Infinite Optimization For a Problem Of Time-Minimal Control, *Journal of Computational Technologies* **5** (1999), S. 67–81

ANDREAS NOLTE, RAINER SCHRADER: Coloring in Sublinear Time, *submitted to: Discrete Applied Mathematics* (1999)

ANDREAS NOLTE, RAINER SCHRADER: Simulated Annealing and its Problems to color graphs, *submitted to Combinatorics, Probability & Computing* (1999)

ANDREAS NOLTE, RAINER SCHRADER: A Note on the Finite Time Behaviour of Simulated Annealing, *Mathematics of OR* **3** (2000), S. 476–484

MARKUS PETER, GEORG WAMBACH: N-extendible posets, and how to minimize total weighted completion time, *Discrete Applied Mathematics* **1-3** (2000), S. 157–167

STEFAN PICKL: Convex Games and Feasible Sets in Control Theory, *accepted for publication: Mathematical Methods of Operations Research* (2000)

STEFAN PICKL: Optimization of the TEM Model – Co-Funding and Joint International Emissions Trading, *OR Proceedings 2000* (2000)

STEFAN PICKL, GERHARD-WILHELM WEBER: An Algorithmic Approach by Linear Programming Problems In Generalized Semi-Infinite Optimization, *Journal of Computational Technologies* **3** (2000), S. 62-82

STEFAN PICKL, JÜRGEN SCHEFFRAN: Control and game theoretic assesment of climate change – options for Joint-Implementation, *International Conference on Transition to advanced Market Institutions and Economies* (2000)

STEFAN PICKL: Controllability via an Approximation Problem, *Proceedings of the 3rd MATHMOD VIENNA 2000* (2000)

STEPHAN ROSSWOG, PETER WAGNER: "Car-SPH": A Lagrangian Particle Scheme fot the Solution of the Macroscopic Traffic Flow Equations, *Traffic and Granular Flow '99* (1999)

Preprints

00-386 ULRICH BLASUM, WINFRIED HOCHSTÄTTLER: Application of the Branch and Cut Method to the Vehicle Routing Problem

99-342 ULRICH BLASUM, WINFRIED HOCHSTÄTTLER, PETER OERTEL: Steiner-Diagrams

00-393 NILS EISSFELDT, STEPHAN ROSSWOG, PETER WAGNER: Microscopic traffic simulation tools and their use for emission calculations

00-395 THOMAS EPPING, WINFRIED HOCHSTÄTTLER, PETER OERTEL: A Paint Shop Problem for Words

99-349 ANDREAS ERDMANN, ANDREAS NOLTE: A Note on Constrained Shortest Path Algorithms in a Column Generation Framework

99-350 ANDREAS ERDMANN, ANDREAS NOLTE, JONAS RATHER, RAINER SCHRADER: A Fast Solution Strategy for the Crew Pairing Problem

99-363 ULRICH FAIGLE, WALTER KERN, JEROEN KUIPERS: On the computation of the nucleolus of a cooperative game

99-362 ULRICH FAIGLE, WALTER KERN: An algebraic framework for the greedy algorithm with applications to the core and Weber set of cooperative games

99-361 ULRICH FAIGLE, WALTER KERN, D. PAULUSMA: Note on the computational complexity of least core concepts for min-cost spanning tree games

00-385 WINFRIED HOCHSTÄTTLER, PETER OERTEL: The 5-Star-Hub-Problem is NP-complete

99-365 WINFRIED HOCHSTÄTTLER, ALEXANDER SCHLIEP: *Developing Gato and CATBox with Python: Teaching graph algorithms through visualization and experimentation*

99-364 ALEXANDER SCHLIEP, FILIPPO CASTIGLIONE: *Detecting in-born metabolic disease in newborns by pattern classification of NMR spectra*

Vorträge

S. FEKETE: *Mehrdimensionales Packen: Probleme, Geometrie, Kombinatorik, exakte Algorithmen*, Eingeladener Vortrag, Universität Düsseldorf 08.01.1999

A. SCHLIEP: *Screening large clone libraries efficiently*, PMMB VI – Program in Mathematics and Molecular Biology, Santa Fe, NM, USA, 09.01.-14.01.1999

U. FAIGLE: *Computing the nucleolus of a cooperative game*, Oberwolfach, 14.01.1999

S. FEKETE: *Simplicity and Hardness of the Geometric Maximum Traveling Salesman Problem*, Symposium on Discrete Algorithms, Baltimore, USA 19.01.1999

S. FEKETE: *Algorithms for More-Dimensional Packing*, Eingeladener Vortrag, Universität Eindhoven, Niederlande 23.02.1999

S. PICKL: *The τ -value as control parameter - an equivalence theorem*, International Conference on Optimization (SIGOPT '99), Trier 22.03.-24.03.1999

M. HAYER: *Test Sets for Vertex Cover Problems*, 6th Twente Workshop on Graphs and Combinatorial Optimization, 26.-28.05, 1999

P. OERTEL: *Steiner Diagrams*, 6th Twente Workshop on Graphs and Combinatorial Optimization, 26.-28.05.1999

S. PICKL: *The τ -value as a control parameter. Modeling and Analysis of a Joint-Implementation Program*, International Conference on Mathematical Modelling and Optimization 1999, Darmstadt 07.07.-09.07.1999

A. SCHLIEP: *Learning Hidden Markov Model Topology*, ISMB'99 – Intelligent Systems in Molecular Biology, Heidelberg, 06.08.-10.08.1999

S. PICKL, G.-W. WEBER: *Theoretical and Practical Relations between Discrete Optimization, Nonlinear Optimization and Control*, Symposium über Operations Research 1999, Sektion: Optimal Control, Magdeburg 01.09. - 3.9.1999

S. PICKL, G.-W. WEBER: *Analysis and Optimization of the TEM-Model with Polytopes*, Symposium über

Operations Research 1999, Sektion: Combinatorial and Discrete Optimization, Magdeburg 01.09.-03.09.1999

S. PICKL: *Using Polytopes for Detecting Stability in the TEM Model*, Deutsche Mathematiker Vereinigung Jahrestagung DMV 1999, Mainz 06.09.-10.9.1999

P. OERTEL: *Steiner Diagrams and k-Star-Hubs*, FRICO 99, Konstanz, 01.-02.10.1999

A. SCHLIEP: *Learning Hidden Markov Model Topology*, GCB'99 – German Conference on Bioinformatics, Hannover, 04.10.-06.10.1999

A. SCHLIEP: *Clustering Protein Sequences – Structure Prediction by Transitive Homology*, GCB'99 – German Conference on Bioinformatics, Hannover, 04.10.-06.10.1999

S. PICKL: *Aktivierende Lehr- und Lernformen*, Posterpräsentation, 6. Werkstattgespräch zur Verbesserung der Lehre, Darmstadt, 22.10.1999

S. PICKL: *Verantwortliche Energieversorgung für die Zukunft — Kooperative Lösung technikbedingter Konflikte am Beispiel von Joint Implementation*, Tagung „Wissenschaft, Technik und Ethik“, München, 08.-09.11.1999

A. SCHLIEP: *Learning Hidden Markov Model Topology for Sequence Analysis*, In silico Biology: Sequence, Structure and Function. The Second International Georgia Tech Conference on Bioinformatics, Atlanta, GA, USA, 11.11.-14.11.1999

A. SCHLIEP: *Clustering Protein Sequences*, TIGR-Computational Genomics Conference III, Baltimore, MD, USA, 18.11.-21.11.1999

A. SCHLIEP: *Screening large clone libraries efficiently*, NCBI - National Center for Biotechnology Information, NIH, Bethesda, MD, USA, 23.11.1999

S. PICKL: *Identification, Analysis and Environmental Protection with the TEM-Model - Simulation of an Economic Joint-Implementation Program*, International Workshop of the GOR Working Group „OR and Environmental Protection“, Braunschweig 25.11.-26.11.1999

- U. FAIGLE: *Optimierung und Spieltheorie*, Technische Universität Darmstadt, 01.12.1999
- M. METZLER: *3V – Verteilte Verkehrssimulation und Visualisierung*, Praxis der Datenverarbeitung, RRZK, Universität zu Köln, 08.12.1999
- S. HASSELBERG: *Bestimmung von kürzesten Wegen mit Hilfe einer Baumheuristik*, Oberseminarvortrag, MI, Universität zu Köln, 21.01.2000
- M. METZLER: *Abschlußbericht 3V*, Abschlußveranstaltung der GTB West, Jülich, 31.01./01.02.2000
- S. PICKL: *Controllability of the Time-Discrete TEM-Model via the Solution of an Approximation Problem*, 3rd MATHMOD VIENNA - 3rd IMACS Symposium on Mathematical Modelling, Wien 02.02.-04.02.2000
- S. PICKL: *A New Algorithm for the Nonlinear Time-Discrete TEM-Model Applying Game Theory with Discrete and Dynamic Methods*, Internationaler Workshop der GOR-Arbeitsgruppe „OR im Umweltschutz“ New Approaches with Dynamic Games and Operations Research Methods, Darmstadt 13.03.-14.03.2000
- W. KRABS, S. PICKL: *On Norm-Minimal Local Controllability of Time-Discrete Dynamical Systems*, 10. Workshop der GOR-Arbeitsgruppe „Entscheidungstheorie und -praxis“, Frankfurt am Main 16.03.-18.03.2000
- S. PICKL: *A Dynamic Model for a Joint-Implementation Program - The Kyoto Game*, 64. Physikertagung der Deutschen Physikalischen Gesellschaft, Dresden 20.03.-24.03.2000
- R. SCHRADER: *Mikroskopische Simulation, dynamische Routensuche und -umlegung*, BMBF-Workshop ILUMASS beim TÜV Rheinland, Köln, 30.03.2000
- W. KRABS, S. PICKL: *Optimization of an n-Person Game under Linear Side Conditions. A Theoretic and Numerical Approach to an Actual Eco-Environmental Problem concerning Kyoto Protocol*, Seventh Viennese Workshop on Optimal Control, Dynamic Games and Nonlinear Dynamics, Wien 24.05.-26.05.2000
- S. PICKL: *Existence of Optimal Control Parameters in Nonlinear Time-Discrete Dynamic Systems. A Topological and Numerical Approach concerning Kyoto Protocol*, Seventh Viennese Workshop on Optimal Control, Wien 24.5.-26.5.2000
- N. EISSFELDT: *Dynamic simulation of traffic and its environmental impacts*, 9th IFAC Symposium Control in Transportation Systems, Braunschweig, 13.06.-15.06.2000
- G.W. WEBER: *Semi-infinite optimization, optimal control and some relations with interior point methods*, Workshop on Interior Point Methods IPM-2000 Budapest, 07.2000
- G.W. WEBER: *On generalized semi-infinite optimization*, European Conference on Operational Research, Budapest, 07.2000
- S. PICKL: *Optimierung von CO₂-Minderungsaktivitäten mithilfe von TEMPI*, DFG-Workshop „Human Dimensions of Global Change“, Schloss Wendgraben, 13.-15.07.2000
- A. SCHLIEP: *Clustering Protein Sequences – Structure Prediction by transitive homology*, ISMB'00 – Intelligent Systems in Molecular Biology, San Diego, CA, USA, 19.08.-23.08.2000
- G.W. WEBER: *On generalized semi-infinite optimization: relations to discrete structures and optimal control*, International Symposium on Mathematical Programming, Atlanta (USA), 08.2000
- TH. EPPING: *Ein „paint shop“-Problem für Wörter*, FRICO 2000, Aachen, 01.09.-02.09.2000
- CH. HAGEMEIERS: *Tail Assignment bei Charterfluggesellschaften unter Beachtung von Wartungsereignissen*, FRICO 2000, Aachen, 01.09.-02.09.2000
- CH. MUES: *Ein Pickup & Delivery-Problem mit Umlademöglichkeit*, FRICO 2000, Aachen, 01.09.-02.09.2000
- D. RÄBIGER: *LP basierte Lösungsverfahren für das Fleet Assignment Problem*, FRICO 2000, Aachen, 01.09.-02.09.2000
- S. PICKL: *Optimization of the TEM Model. Co-Funding and Joint International Emissions Trading (IET)*, Symposium über Operations Research SOR 2000, 09.09.-12.09.2000
- S. PICKL: *The TEM Model — Simulation and optimization of CO₂ reduction programs*, Vortrag anl. der Verleihung des Dissertationspreises 1998/99 der GOR, Dresden, 10.09.2000
- S. PICKL: *Charakterisierung und Analyse kombinatorischer Strukturen zulässiger Mengen mithilfe von*

Maximumfunktionen und des Clarke'schen Subdifferentials, DMV-Jahrestagung, Dresden, 18.09.-22.09.2000

G.W. WEBER: *Generalized semi-infinite optimization: continuous and discrete*, French-German-Italian Conference on Optimization in Montpellier (Frankreich), 09.2000

G.W. WEBER: *Some foundations and applications of generalized semi-infinite optimization*, Symposium on Operations Research OR 2000 – Jahrestagung der GOR in Dresden, 09.2000

G.W. WEBER: *Generalized semi-infinite optimization: relations with optimal control and discrete optimization*, Jahrestagung 2000 der DMV in Dresden, 09.2000

G.W. WEBER: *Verallgemeinerte semi-infinite Optimierung und optimale Steuerung: Theorie, Methoden und Anwendungen – aus stetiger und diskreter Sicht*, Besuch der Technischen Universität Clausthal, Institut für Mathematik, 09.2000

L. KADERALI, A. SCHLIEP: *An algorithm for selecting target specific probes for DNA arrays (Posterpräsentation)*, GCB'00 – German Conference on Bioinformatics, Heidelberg, 07.10.-10.10.2000

P. OERTEL: *Routing mit Umladen*, Siemens Workshop, Würzburg, 22.10.-24.10.2000

N. EISSFELDT: *Fragestellungen dynamisch-mikroskopischer Verkehrssimulation in der Anwendung*, Workshop „Discrete Optimization — Structure and Stability of Dynamical Systems“, Köln, 26.10.2000

CH. HAGEMEIERS: *Dienstplanung im ÖPNV*, Workshop „Discrete Optimization — Structure and Stability of Dynamical Systems“, Köln, 26.10.2000

S. PICKL: *Charakterisierung und Analyse kombinatorischer Strukturen zulässiger Mengen mithilfe von Maximumfunktionen und des Clarke'schen Subdifferentials*, Workshop „Discrete Optimization — Structure and Stability of Dynamical Systems“, Köln, 26.10.2000

G.W. WEBER: *On generalized semi-infinite optimization einer Session*, Besuch der Technischen Universität Clausthal, Institut für Mathematik, 10.2000

G.W. WEBER: *Approaches to mechanics from discrete and nonlinear optimization and control*, Discrete Modeling and Discrete Algorithms in Continuum Mechanics in Braunschweig, 10.2000

P. OERTEL: *Algorithmen für Speditionsroutingprobleme mit Umlademöglichkeit*, 6. Magdeburger Logistiktagung, Magdeburg, 16.11.-17.11.2000

A. SCHLIEP: *Developing Gato and CATBox with Python: Teaching graph algorithms through visualization and experimentation*, Multimedia Tools for Communicating Mathematics, Lissabon, Portugal, 23.11.-25.11.2000

Weitere Teilnahmen und Forschungsbesuche

S. PICKL: *Zertifiziertes Habilitandenprogramm „Studieren Lehren Forschen“*, TU Darmstadt, 1999

CH. HAGEMEIERS: *Forschungsaufenthalt bei Prof. Dr. Prömel*, Humboldt Universität zu Berlin, Berlin, Juni 1999

S. PICKL: *Forschungsaufenthalt bei Prof. Dr. M. Hasler*, EPFL Lausanne, Juli 1999

S. PICKL: *Einladung zur Late Summer School Facets of the Polytope World*, Institut für Theoretische Informatik, ETH Zürich (Prof. Dr. E. Welzl, Prof. Dr. J. Richter-Gebert), Zürich, Schweiz, 13.09.-16.09.1999

S. PICKL: *Forschungsaufenthalt an der Universität Tilburg*, Department of Econometrics (Seminar Prof. Dr. S. Tijs), Tilburg, 21.02.-25.02.2000

A. SCHLIEP, L. KADERALI, A. SCHÖNHUTH: *Sommerschule: Mathematical Methods for Protein Structure Analysis and Design*, Martina Franca, Italien, 9.07.-15.07.2000

P. OERTEL, M. PETER : *1. Max-Planck Advanced Course on the Foundations of Computer Science*, 31.08.-04.09.2000,

S. PICKL: *Teilnahme an der Sommerschule Bioinformatik*, Universität Bielefeld, Technische Fakultät (Prof. Dr. R. Giegerich), Bielefeld, 12.09.-16.09.2000

S. PICKL: *Teilnahme an der Herbstschule Bioinformatik*, FU Berlin, Fakultät für Informatik (Prof. Dr. G. Rode), Berlin, 6.11.-10.11.2000

Fertiggestellte Dissertationen

U. BLASUM: *Anwendung des Branch & Cut Verfahrens auf das kapazitierte Vehicle Routing Problem*

F. BROCKNERS: *Flußsteuerung für verlässliche Multicastkommunikation im Internet*

A. ERDMANN: *Combinatorial Optimization Problems arising in Airline Industry*

CH. GAWRON: *Simulation-Based Traffic Assignment – Computing User Equilibria in Large Street Networks*

S. HASSELBERG: *Some results on heuristical algorithms for shortest path problems in large road networks*

M. HORNEFFER: *Methoden zur Leistungsbewertung von Internet-Zugangspunkten*

B. KNAB: *Erweiterungen von Hidden-Markov-Modellen zur Analyse ökonomischer Zeitreihen*

A. NOLTEMEIER: *Modelle und Lösungsverfahren zur Generierung von Flugplänen*

I. WEBER: *SVD-Clusterung von Bauspar-Zeitreihen*

K. WEINBRECHT: *Kontinuierliche Standortprobleme in Polygonen*

R. WIMMERSHOFF: *Untersuchungen zur parallelen, objektorientierten Verkehrssimulation*

Vorgelegte Diplomarbeiten

I. FEINHOLDT: *Ein praktischer Algorithmus für das Fussballproblem*

D. GROTH: *Sensitivitätsanalyse für kürzeste Wege auf Straßengraphen*

CH. HAGEMEIER: *Branch & Price Verfahren für das Duty Scheduling Problem*

S. JANZ: *Mikroskopische Minimalmodelle des Straßenverkehrs*

CH. MUES: *Ein Pickup & Delivery-Problem mit Umladung von Gütern*

D. RÄBIGER: *LP basierte Lösungsverfahren für das Fleet Assignment Problem*

J. RATHERT: *Branch-and-Price Verfahren für das Crew-Scheduling*

A. VILDOSOLA: *Berechnung des Benutzeroptimums in Verkehrsnetzen mit Hilfe des Frank-Wolfe-Algorithmus*

Anhang E

Mitarbeiter 1999/2000

Lehrstuhlinhaber

Prof. Dr. Ulrich Faigle	0221/470-6030
Prof. Dr. Rainer Schrader	0221/470-6030

Sekretariat

Claudia Teuner	0221/470-6030
----------------	---------------

Mitarbeiter¹

Kamel Ben Khalifa	6020	Silja Meyer–Nieberg	6019
Thomas Chevalier	6023	Christopher Mues	6050
Nils Eissfeldt	6026	Martin Olschewski	6018
Petra Fakler	6023	Markus Peter	6022
Bernhard Fuchs	6038	Dr. Stefan Pickl	6024
Jürgen Gräfe	6010	Dirk Rübiger	6047
Christian Hagemeyer	6048	Urs Roth	6019
Matthias Hayer	6011	Alexander Schönhuth	6003
Lars Kaderali	6003	Vanessa Stausberg	
Markus Kloock	6038	Bernd Wichern	6017

Mitarbeiter der Verkehrsgruppe des DLR ²

Georg Hertkorn	6043 (030/67055-237)
Dr. Peter Wagner	6043 (030/67055-232)

¹(0221) 470-xxxx. Die email-Adressen der Mitarbeiter setzen sich so zusammen: nachname@zpr.uni-koeln.de.

²(0221) 470-xxxx. Die email-Adressen setzen sich ebenso zusammen, wie die der Institutsmitarbeiter. In Klammern ist die Telefonnummer beim DLR angegeben.

Studentische Hilfskräfte

Ramazan Buzdemir	6021	Philipp Messer	6021
Natalia Gayeva	6019	Torsten Pattberg	6011
Christian Gießelbach	6021	Peter Pipenbacher	6021
Achim Gädke	6021	Christian Rössel	6043
René Kermis	6001	Olaf Schumann	6021
Oliver Klein	6000	Leonid Schwab	6000
Markus Kleinhans	6021	Linda Thorarinsdottir	6021
Philipp Melchior	6043	Olaf Wendisch	6021

Ehemalige Mitarbeiter

Dr. Ulrich Blasum	Dr. Markus Metzler
Dr. Rolf Böning	Dr. Andreas Nolte
Dr. Frank Brockners	Dr. Anja Noltemeier
Thomas Epping	Dr. Peter Oertel
Dr. Andreas Erdmann	Thomas Olschewski
Prof. Dr. Sándor Fekete	PD Dr. Gerhard Weber
Dr. Stephan Hasselberg	Dr. Iris Weber
Prof. Dr. Winfried Hochstättler	Dr. Karin Weinbrecht
Holger Klein	Katja Wolf
Dr. Bernhard Knab	

Leitung / Sekretariat

Prof. Dr. Ulrich Faigle



Prof. Dr. Rainer Schrader



Claudia Teuner

Mitarbeiter

Kamel Ben Khalifa



Filippo Castiglione



Thomas Chevalier



Nils Eissfeldt



Thomas Epping



Jürgen Gräfe



Christian Hagemeyer



Matthias Hayer



PD Dr. Winfried Hochstättler



Lars Kaderali



Markus Kloock



Christopher Mues



Martin Olschewski



Markus Peter



Dr. Stefan Pickl



Dirk Rübiger



Urs Roth



Alexander Schliep



Alexander Schönhuth



Vanessa Stausberg



Bernd Wichern



Georg Hertkorn

Studentische Hilfskräfte



Ramazan Buzdemir



Natalia Gayeva



Christian Giebelbach



Achim Gädke



René Kermis



Oliver Klein



Markus Kleinhans



Philipp Melchior



Philipp Messer



Torsten Pattberg



Peter Pipenbacher



Christian Rössel



Olaf Schumann



Leonid Schwab



Linda Thorarinsdottir



Hartmut Wahl



Olaf Wendisch

Hutmoden



Iris Weber und Bernhard Knab



Karin Weinbrecht und Ulrich Blasum



Anja Noltemeier