

ANGEWANDTE MATHEMATIK UND INFORMATIK UNIVERSITÄT ZU KÖLN

Report No. 97.295

Mesoskopisches Simulationsmodell zur Kollektivfortschreibung

by

B. Knab, R. Schrader, I. Weber, K. Weinbrecht, B. Wichern

1997

Ausarbeitung eines auf der Fachtagung „Kollektivanalyse und -fortschreibung“ am 16.10.1997 in
Stuttgart gehaltenen Vortrages

Zentrum für Paralleles Rechnen
Universität zu Köln
Albertus-Magnus-Platz
D-50923 Köln

1991 Mathematics Subject Classification: 62H30, 90A09, 90A19, 90A20, 90B10, 92G30
Keywords: cluster analysis, single link, complete link, network flows

Mesoskopisches Simulationsmodell zur Kollektivfortschreibung

Bernhard Knab, Rainer Schrader, Iris Weber, Karin Weinbrecht, Bernd Wichern
Zentrum für Paralleles Rechnen, Universität zu Köln

Einleitung

Kollektivanalysen und darauf aufbauende Prognosen sind seit langem ein wichtiger Beitrag der Bausparmathematik zu Fragen der Liquiditätsplanung, der Produktpflege und der Produktentwicklung. Im Rahmen einer langjährigen Kooperation zwischen den Landesbausparkassen und dem Zentrum für Paralleles Rechnen wurden daher Bausparmodelle entwickelt, die der Analyse des Verhaltens der Bausparer und der Vorhersage ihres zukünftigen Verhaltens dienen. Eine Weiterentwicklung dieser Modellansätze soll im folgenden vorgestellt werden.

Analysen und Prognosen

Der Ablauf eines Bausparvertrages teilt sich in der Regel in drei Phasen: die Sparphase, die Zuteilung und die Darlehensphase. Diese drei Phasen lassen sich getrennt voneinander analysieren, wobei in einem zweiten Schritt Korrelationen zwischen den Verhaltensmustern in den einzelnen Phasen zu untersuchen sind. Von besonderem Interesse ist die Sparphase, da in diesem Abschnitt eines Bausparvertrages — anders als während der Rückführung des Baudarlehens in der Darlehensphase — der Bausparer weitgehend frei ist und individuelle Sparziele verfolgen kann.

Im Hinblick auf ein Bausparkollektiv gibt es grundsätzlich zwei für die Produktsteuerung wichtige Fragestellungen: Wie hat sich das Kollektiv in der Vergangenheit entwickelt (Kollektivanalysen), und wie wird sich das Kollektiv weiterentwickeln (Prognosen). Dabei dienen die Analysen der eher kurzfristigen Liquiditätsplanung und als Grundlage für die Planung von Marketingstrategien, während Prognosen für die mittel- und langfristige Liquiditätsplanung sowie für die Produktentwicklung eingesetzt werden.

Bisherige Prognosemodelle

Für kurz- und mittelfristige Vorhersagen ist der Einfluß des bestehenden Kollektives von vorherrschender Bedeutung. Daher müssen Prognosemodelle, die verlässliche Aussagen treffen sollen, eine getreue Abbildung des Bestandes liefern. Für längere Prognosehorizonte ist dagegen die Fortschreibung des Kollektives in Verbindung mit exogenen Annahmen ausschlaggebend.

In der Vergangenheit dienten ein Schichtenmodell und ein Mikro-Simulations-Modell zur Erstellung von Prognosen. Im Schichtenmodell wird ein vorgegebenes Bausparkollektiv durch eine kleine Zahl von Gruppen von Verträgen (Schichten) abgebildet, die jeweils ein homogenes Verhalten zeigen. Die Schichten werden so gewählt, daß sie die von den Bausparkassen beobachteten typischen Verhaltensmuster abdecken, etwa Regelsparer, die eine konstante Sparleistung erbringen, Soforteinzahler oder Zuzahler. Die Anteile dieser Schichten am bestehenden Kollektiv werden mittels eines nichtlinearen Optimierungsansatzes so bestimmt, daß jährlich erhobene Kollektivgrößen möglichst genau abgebildet werden. In der Simulation werden dann die Neugeschäftsanteile der Schichten deterministisch fortgeschrieben. Dieses Schichtenmodell ist für

langfristige Prognosen geeignet, die Abbildung nichtkonstanten Sparerverhaltens und flexiblerer Tarife ist jedoch problematisch. Seine Schwächen in der kurz- und mittelfristigen Prognose resultieren überwiegend aus der groben Abbildung des aktuellen Kollektives, auf das in der Simulation aufgesetzt wird. So kann z. B. nur eindeutig gesagt werden, welchen Anteil zwei Gruppen insgesamt bekommen, die sich in den ersten Jahren, in denen sie bisher gelaufen sind, völlig identisch verhalten. Es kann aber nicht eindeutig gesagt werden, wie sich dieser Anteil auf die beiden Gruppen aufteilt.

Für mittel- und kurzfristige Prognosen wurde ein Mikro-Simulations-Modell entwickelt, das auf den jahres- oder quartalsweise erhobenen Einzelverträgen einer Bausparkasse basiert und diese Verträge stochastisch fortschreibt. Ein Vorteil dieser Vorgehensweise liegt in der exakten Abbildung des Bestandes zu Beginn der Simulation. Eine große Schwierigkeit bei der Anwendung dieses Modells besteht in seiner hohen Komplexität, insbesondere in der geeigneten Wahl der Parameter.

Ein neuer Modellansatz

In dem nun entwickelten neuen Modellansatz sollen die Vorteile der beiden bisher entwickelten Modelle kombiniert und ihre Nachteile eliminiert werden. Dabei werden weiterhin die schon für das Mikro-Simulations-Modell erhobenen Daten zur Abbildung des Bestandes zugrundegelegt. Es handelt sich dabei i. a. um 500.000 bis 3 Mio. Einzelverträge pro Bausparkasse mit jeweils 49 Kontoeinträgen wie Vertragsnummer, Vertragssumme, Guthaben, Darlehensstand, Spargeldzugang, Guthabensauszahlung etc. für jeden Zeitraum.

Wie schon dem älteren Schichtenmodell liegt auch dem neuen Modellansatz die Idee zugrunde, daß Verträge mit ähnlichem Muster im Sparverhalten für die Prognose zu Gruppen zusammengefaßt werden können (vgl. Abbildung 1).

Idee: Gruppen ähnlichen Verhaltens für die Prognose zusammenfassen

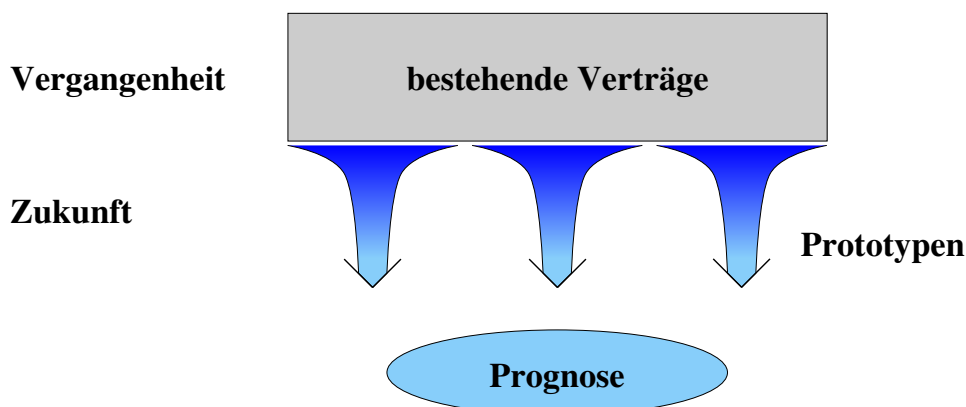


Abbildung 1: Modellansatz

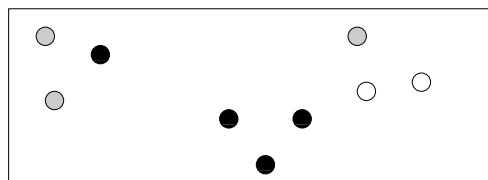
Anders als im älteren Modell werden diese Gruppen aber auf der Basis der Einzelverträge definiert, und ihre Anteile werden ebenfalls auf der Basis der Einzelverträge festgelegt, so daß der

Bestand exakt abgebildet wird. Bestimmt werden diese Gruppen außer durch ihr Sparverhalten auch noch durch ihre gesamte Vertragssumme zu einem bestimmten Abschlußzeitpunkt.

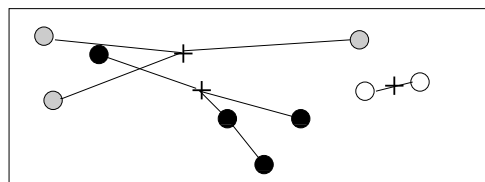
Einsatz von Clusterverfahren

Zunächst werden nur diejenigen Verträge betrachtet, deren Sparphase vollständig bekannt oder besonders lang ist. Das Zusammenfassen dieser Verträge zu verschiedenen Gruppen erfolgt mit Hilfe von Clusteralgorithmen. Dazu wird über den jährlichen Spargeldeingang der Verträge und weitere Merkmale ein Abstandsmaß definiert. Ein Cluster, also eine Gruppe von Verträgen, wird durch seinen über die gleichen Merkmale wie das Abstandsmaß berechneten Schwerpunkt repräsentiert. Dabei muß das verwendete Abstandsmaß verträglich mit der Berechnung des Schwerpunktes sein, außerdem wird jeder Vertrag mit seiner Vertragssumme gewichtet. Ziel der Clusteranalyse ist es, die Cluster so zu bestimmen, daß die Abstände der Verträge zu ihren Clusterschwerpunkten möglichst klein werden und somit möglichst homogene Sparergruppen entstehen. Die Schwerpunkte dieser berechneten Cluster bestimmen das Verhalten der Prototypen für die Simulation.

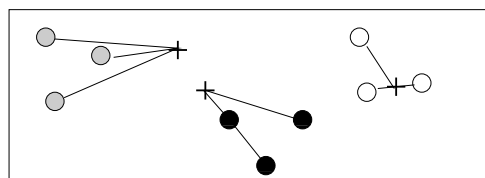
Wähle zufällige Startpartition



Berechne Zentralpunkte



Bilde Minimaldistanzpartition



Endpartition

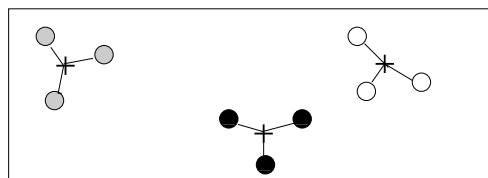


Abbildung 2: Minimaldistanzverfahren

Mathematisch läßt sich das Clusterproblem für das hier verwendete Clusterverfahren, die Centroidmethode, folgendermaßen als Minimierungsaufgabe formulieren:

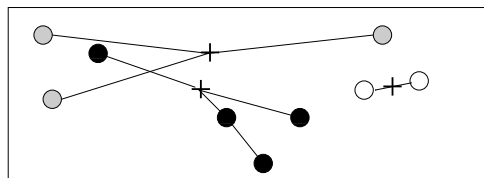
Gesucht ist eine Verteilung von m Objekten auf K Cluster derart, daß die Summe der Abstände der Objekte A_{jk} aller Cluster zu ihrem jeweiligen Mittelwertvektor Z_k minimal wird:

$$\text{Minimiere } \sum_{k=1}^K \sum_{j=1}^{n(k)} d(A_{jk}, Z_k) \quad \text{mit } \sum_{k=1}^K n(k) = m.$$

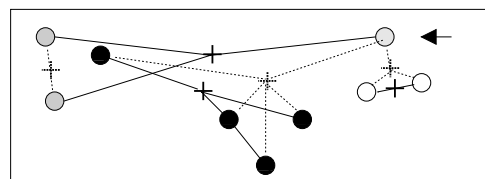
Dieses Problem ist NP-schwer, d. h. es ist kein deterministischer Algorithmus bekannt, der das Problem effizient (in polynomieller Zeit) exakt löst. Man ist deshalb auf Heuristiken angewiesen, die eine gute Lösung — im allgemeinen ein lokales Minimum — in angemessener Zeit liefern. Zwei von uns eingesetzte Methoden sind das Minimaldistanzverfahren und das Austauschverfahren, wobei sich in der Praxis die Kombination der beiden genannten Verfahren bewährt hat.

Beim Minimaldistanzverfahren werden abwechselnd alle Objekte ihrem nächsten Zentralpunkt zugeordnet und dann die Zentralpunkte neu berechnet (vgl. Abbildung 2).

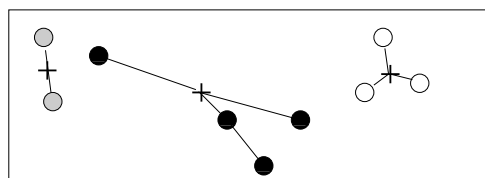
Wähle zufällige Startpartition



Teste ein Objekt in allen Clustern



Update der Zentralpunkte



Endpartition

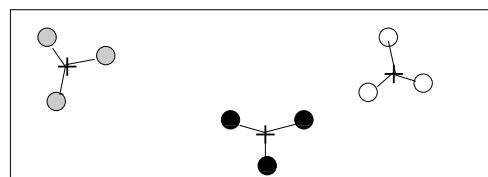


Abbildung 3: Austauschverfahren

Beim Austauschverfahren wird für ein bestimmtes Objekt geprüft, ob durch Zuordnung dieses Objekts zu einem anderen Cluster die Summe der Abstände zum jeweiligen Mittelpunkt für alle Objekte verringert werden kann. Ist dies der Fall, wird dieses eine Objekt umsortiert und der Schritt erneut durchgeführt (vgl. Abbildung 3).

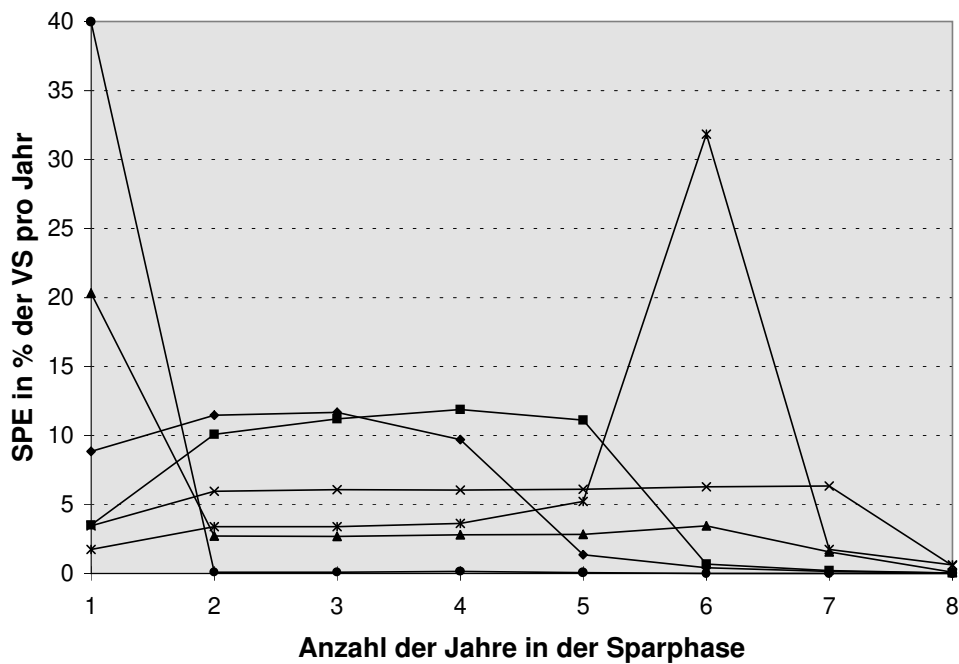


Abbildung 4: Prototypen für das Ansparverhalten

Auf die beschriebene Weise lassen sich aus den Verträgen mit vollständig bekannter Sparphase Prototypen für das Ansparverhalten gewinnen. Obwohl dieses Verfahren automatisch und ohne Zusatzinformationen abläuft, stellt sich heraus, daß sich unter den so erhaltenen Prototypen auch die von den Bausparkassen empirisch beobachteten und im Schichtenmodell berücksichtigten Sparer wie Regelsparer, Soforteinzahler, Zuzahler etc. wiederfinden (vgl. Abbildung 4).

Vervollständigung der Prototypen

Um das Verhalten der Sparer zu vervollständigen, müssen außer dem Ansparverhalten noch Abruf- und Tilgungsverhalten definiert werden. Unter Abrufverhalten sind die Zeitpunkte und jeweiligen Höhen der Inanspruchnahme der Guthabens- und Darlehensauszahlungen zu verstehen. Die durchgeführten Untersuchungen rechtfertigen die Auszahlung des gesamten Guthabens bei Zuteilung und des gesamten Darlehens mit einem festen Vershub bezüglich der Zuteilung. Clusterungen der Darlehensphase legen zwei typische Verhaltensweisen für die Tilgung nahe: Zum einen tritt wie zu erwarten tarifliche Tilgung sehr oft auf, zum andern lassen sich die übrigen Verhaltensweisen vereinfachend so darstellen, daß zunächst nach Tarif getilgt und dann nach einer gewissen Zeit durch eine einmalige Sondertilgung das Restdarlehen vollständig abgelöst wird.

Eine besondere Behandlung erfordern Verträge, die nach dem Ende ihrer Sparphase nicht oder nicht sofort in die Zuteilungs- und anschließend in die Darlehensphase übergehen oder die sich schon während ihrer Sparphase durch ein gewisses Sonderverhalten auszeichnen. Es treten zwei Arten von Sonderverhalten auf: Für die eine — zu ihr zählen Kündigung, Fortsetzung und Darlehensverzicht — werden eigene Gruppen gebildet, deren Verhalten sich aus den für den Gesamtbestand ermittelten Prototypen und ihrem Sonderverhalten zusammensetzt. Mit Hilfe von aus der Vergangenheit gewonnenen Häufigkeiten werden dann für die einzelnen Prototypen Anteile festgelegt, die zu einem bestimmten Zeitpunkt in ein bestimmtes Sonderverhalten dieser

Art abzweigen. Für die andere Art von Sonderverhalten — Erhöhung oder Ermäßigung der Vertragssumme — wurden der Zeitpunkt und der Betrag der Vertragssummenänderung als weitere Merkmale bei der Abstands- und Schwerpunktsberechnung der Clusteranalyse berücksichtigt und so eigene Erhöher- und Ermäßigerprototypen definiert.

Die sich aus diesem Vorgehen ergebenden Kombinationsmöglichkeiten für das Verhalten eines Sparers sind in Abbildung 5 dargestellt.

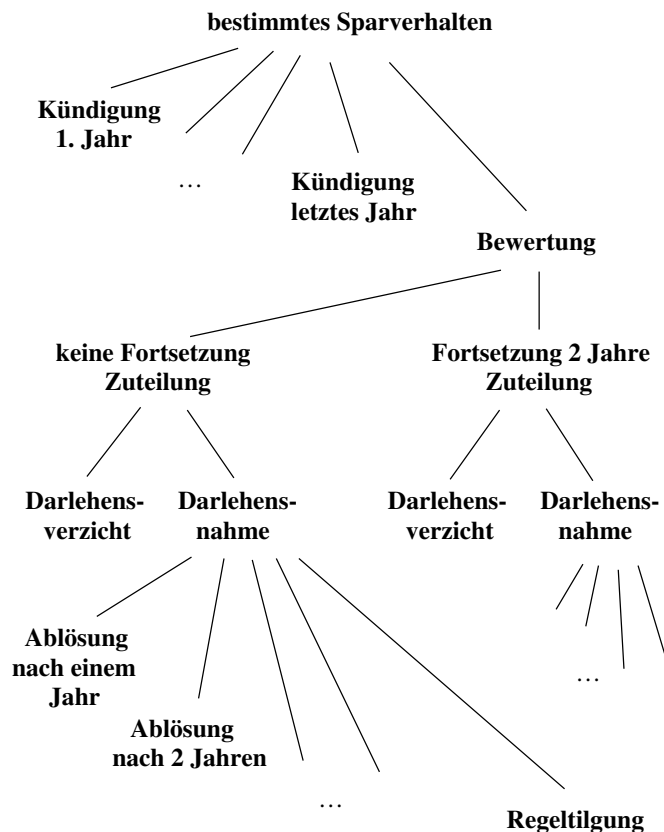


Abbildung 5: Kombinationsmöglichkeiten der Verhaltensweisen

Einsatz eines Netzwerkfluß-Algorithmus

Die bisherige Einschränkung auf Verträge, deren Sparphase vollständig bekannt ist, diente lediglich der Definition prototypischen Verhaltens. Da diese Verträge weitestgehend abgewickelt sind, sind sie für die eigentliche Prognose nur in ihrer Darlehensphase von Bedeutung.

Im nächsten Schritt muß daher festgelegt werden, welchem Prototypen ein im aktuellen Bestand vorhandener Sparer, dessen Sparphase noch nicht abgeschlossen ist, zugeordnet werden soll. Vor allem junge Verträge, die nur über einen kurzen Zeitraum beobachtet werden können, lassen sich in der Regel nicht eindeutig zuordnen (vgl. Abbildung 6). Daher werden für jeden Prototypen Unter- und Obergrenzen der Anteile bzgl. der Vertragssummen ermittelt, die beim Zuordnen ein-

gehalten werden müssen. Über diese Intervalle lassen sich bei der Zuordnung auch die Eigenschaften der Prototypen berücksichtigen, die noch in der Zukunft liegen und deswegen nicht in den Abstand eingehen. So kann beispielsweise verhindert werden, daß ein Prototyp, der in der Zukunft ein sehr extremes Verhalten aufweist, einen sehr hohen Anteil an Verträgen erhält.

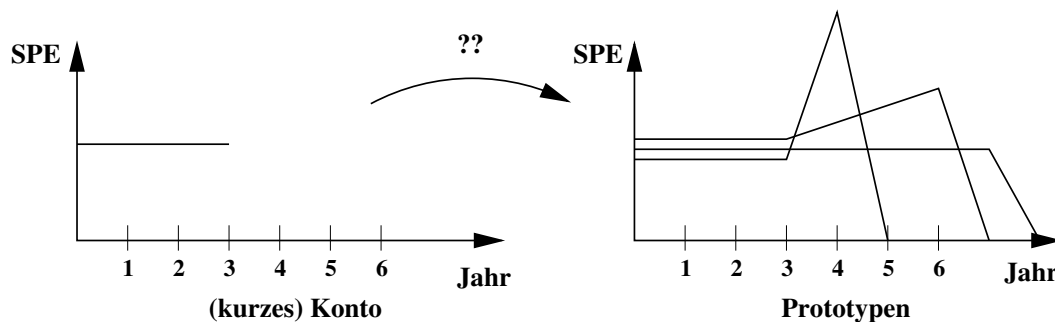


Abbildung 6: Zuordnungsproblematik

Die benötigten Ober- und Unterschranken werden folgendermaßen bestimmt: Zunächst werden für jeden Prototypen diejenigen Verträge, die ihn definiert haben, nach Abschlußjahren getrennt. Dann werden für jedes Abschlußjahr für jeden Prototypen Anteile errechnet. Davon ergibt jeweils das Minimum die Unterschranke, das Maximum die Oberschranke. Dabei dürfen allerdings nur diejenigen Abschlußjahre berücksichtigt werden, die so weit zurückliegen, daß der jeweilige Prototyp tatsächlich eine Chance hatte, seine Sparphase zu beenden.

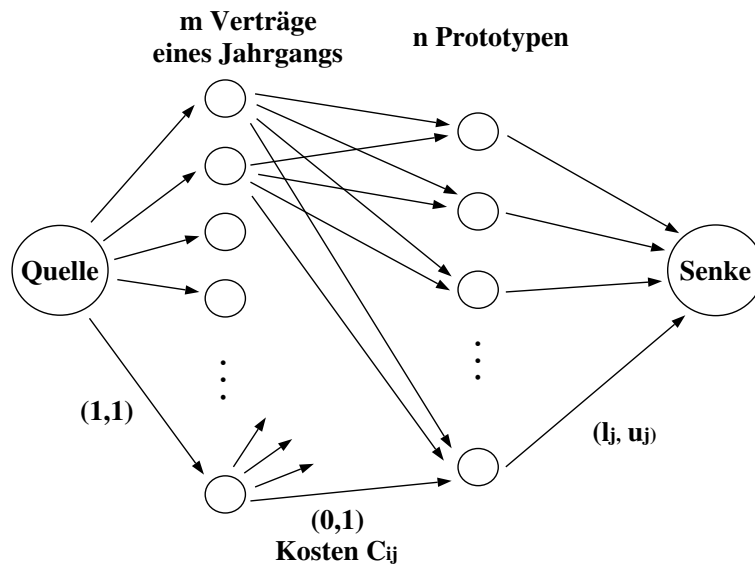
Die Zuordnung selbst erfolgt mit Hilfe eines Minimum-Cost-Flow-Ansatzes (vgl. Ahuja et al., Network Flows) so, daß unter Berücksichtigung dieser Schranken als Nebenbedingungen die Summe der Abstände eines jeden einzusortierenden Vertrages zu seinem Zentralpunkt minimal wird. In der Formulierung als ganzzahliges Flußproblem (vgl. Abbildung 7) sind die Kosten durch die Abstände gegeben. Dieses Problem läßt sich in polynomieller Zeit exakt lösen, und wir setzen hierfür einen von Goldberg entwickelten besonders effizienten Algorithmus (vgl. Goldberg, An Efficient Implementation of a Scaling Minimum-Cost Flow Algorithm) ein.

Zuordnung eines gesamten Bestandes

Innerhalb derjenigen Verträge, die einen Prototypen bestimmen haben, werden nun jeweils die Verhältnisse Kündiger – Nichtkündiger, Fortsetzer – Nichtfortsetzer, Darlehensverzichter – Darlehensnehmer und die Anteile der Kündigungen zu verschiedenen Zeitpunkten ermittelt.

Grundsätzlich darf jeder Vertrag jedem Prototypen zugeordnet werden. Eine Ausnahme bilden Erhöher und Ermäßiger: Ein Vertrag, der bereits seine Vertragssumme verändert hat, soll nicht zu einem Prototypen sortiert werden, der seine Vertragssumme erst in der Zukunft ändern wird. Und umgekehrt soll ein Vertrag, der bisher keine Vertragssummenänderung hatte, auch nicht zu einem Prototypen sortiert werden, der in der Vergangenheit bereits erhöht oder ermäßigt hat. Diese Zuordnungen werden durch künstlich erhöhte Abstände verhindert.

Ebenfalls eine Sonderrolle nimmt der jüngste Jahrgang ein, der zugeordnet werden muß. Hier steht erst sehr wenig Information zur Verfügung, nämlich nur der Spargeldeingang eines einzi-



Minimum-Cost-Flow-Problem:

Finde den Fluß mit den geringsten Gesamtkosten unter Einhaltung der Kapazitäten

$(*,*)$: Unter- und Obergrenze der jeweiligen Kantenkapazität

C_{ij} : Kosten der Zuordnung Vertrag i zu Prototyp j

l_j, u_j : Unter- und Obergrenze des Prototypen j

Abbildung 7: Zuordnung als Minimum-Cost-Flow-Problem

gen Jahres. Es wäre naheliegend, diesen Jahrgang wie das Neugeschäft aufzuteilen, dann würden aber die bereits vorhandenen Bestände an Guthaben und Zinsen nicht exakt getroffen. Dieses Problem kann umgangen werden, indem man diesen Jahrgang nach dem gleichen Verfahren zuordnet wie die älteren Jahrgänge auch, die Unter- und Obergrenzen beim Zuordnen aber auf den gleichen Wert — nämlich den Anteil, den auch das Neugeschäft erhält — setzt. Hier werden die Intervalle also zur Steuerung der Zuordnung eingesetzt.

Die nicht zugeteilten Verträge werden also getrennt nach Abschlußjahren mit Hilfe eines Minimum-Cost-Flow-Algorithmus zu den jeweils in Frage kommenden Prototypen zusortiert. Eine Ausnahme bilden dabei diejenigen Verträge, die bereits als Fortsetzer gekennzeichnet sind: Sie werden zu reinen Bestandsschichten (d. h. diese Schichten bekommen keinen Anteil am Neugeschäft) zusortiert, die jeweils noch eine bestimmte Zeit fortsetzen und dann zugeteilt werden. Ähnlich wird auch mit den bereits zugeteilten Verträgen verfahren: Sie werden nach ihrer (Rest-)Darlehensklasse bzw. ihrem Auszahlungsanspruch in reine Darlehensschichten einsortiert, die entweder bis zum Schluß regelmäßig nach Tarif tilgen oder zu einem bestimmten Zeitpunkt ihr Restdarlehen ablösen oder auf ihren (Rest-)Darlehensanspruch vollständig verzichten. Bei diesen Verträgen spielt das Abschlußdatum keine Rolle.

Simulation

Wurde wie beschrieben der Bestand zu einem bestimmten Zeitpunkt den Prototypen zugeordnet, so müssen noch für die zu simulierenden Zeiträume das Neugeschäftsvolumen und die jeweilige Verteilung auf die einzelnen Prototypen vorgegeben werden. Die Anteile der einzelnen Prototypen am Neugeschäft können beispielsweise aus den jeweiligen Anteilen in der Vergangenheit gemittelt werden. Aggregierte Größen der zu einem Prototypen für ein Abschlußdatum zusortierten Verträge bilden den Startzustand für die Simulation. Danach kann die (wie beim Schichtenmodell deterministische) Bestandsfortschreibung in die Zukunft erfolgen. Der bedeutende Unterschied zum früheren Schichtenmodell besteht darin, daß hier die Vergangenheit nicht nachgerechnet wird, sondern als exakter Bestand eingeht. Für das praktische Rechnen wird ein objektorientiertes Programmpaket zur Simulation von Einzelverträgen und aus Einzelverträgen bestehenden Kollektiven verwendet.

Zusammenfassung und Ausblick

Der beschriebene neue Modellansatz gewährleistet eine exakte Bestandsabbildung bei kurzer Laufzeit und vertretbarer Komplexität und läßt sich somit sinnvoll sowohl für kurz- als auch für mittel- und langfristige Prognosen einsetzen. In weiteren Ausbaustufen können die Automatisierung und die Steuerungsmöglichkeiten erweitert werden, beispielsweise sollen globale Veränderungen der Sparintensität ohne besonderen Aufwand möglich sein, und weitere Differenzierungen im Sparerverhalten wie Tarifwechsel, Zusammenlegung, Teilung etc. können vorgesehen werden.

Literatur

Ahuja, R. K., Magnati, T. L., Orlin, J. B.: *Network Flows*, Prentice Hall, 1993.

Bachem, A., Bettmer, L., Ditzen, S., Windmüller, I.: *Simulationsmodelle für Bausparkkollektive*, in C. Hipp u. a., Hrsg., *Geld, Finanzwirtschaft, Banken und Versicherungen*, Seiten 755 – 761, 1993.

Bachem, A., Fekete, S., Knab, B., Schrader, R., Vannahme, I., Weber, I., Wegener, R., Weinbrecht, K., Wichern, B.: *Analyse großer Datenmengen und Clusteralgorithmen im Bausparwesen*, Technical Report 97-257, Zentrum für Paralleles Rechnen, Universität zu Köln, 1997.

Bachem, A., Korte, B., Schrader, R.: *Mathematische Modelle für Bausparkkollektive*, in Rudolph, B., Wilhelm, J., Hrsg., *Bankpolitik, finanzielle Unternehmensführung und die Theorie der Finanzmärkte*, Seiten 49 – 66, 1988.

Goldberg, A. V.: *An Efficient Implementation of a Scaling Minimum-Cost Flow Algorithm*, Journal of Algorithms 22, 1 – 29, 1997.

Jain, A. K., Dubes, R. C.: *Algorithms for Clustering Data*, Prentice Hall, 1988.

Kellershohn, I.: *Mathematische Simulation von Bausparkkollektiven mit Hilfe von empirischen Verteilungen in monothetischen hierarchischen Kollektivclusterungen*, Dissertation, Köln, 1992.

König, A.: *Neuronale Strukturen zur sichtgestützten Oberflächeninspektion von Objekten in industrieller Umgebung*, Dissertation, Darmstadt, 1995.

Mermagen, R.: *Evaluierung von Clusteralgorithmen für Bausparkkollektive*, Diplomarbeit, Köln, 1995.

Vannahme, I.: *Clusteralgorithmen zur mathematischen Simulation von Bausparkkollektiven*, Dissertation, Köln, 1996.

Anschrift:

Zentrum für Paralleles Rechnen
Universität zu Köln
Albertus-Magnus-Platz
D-50923 Köln