

Ein verbandstheoretisches Modell zu Kreditausfallwahrscheinlichkeiten

Petra Fakler und Rainer Schrader

Wie können Bausparkassen Kreditausfallwahrscheinlichkeiten verlässlich prognostizieren? Diese Frage stellt sich nicht erst seit dem Inkrafttreten von Basel II. Nun wird aber eine externe oder interne Bonitätsbeurteilung zwingend vorausgesetzt, was die Institute vor neue Herausforderungen stellt, da externe Einschätzungen mangels Ratingpraxis in diesem speziell deutschen Geschäft nicht existieren. Eine Lösung kann das verbandstheoretische Implikationenmodell darstellen, so die Autorin, die für ihre Überlegungen den 7. Wissenschaftspreis der privaten Bausparkassen (Werner-Lehmann-Preis) gewonnen hat. (Red.)

Seit dem Jahr 2007 gelten für alle europäischen Kreditinstitute die neuen Eigenkapitalrichtlinien Basel II und die damit verbundene Forderung, Risiken angemessen durch Eigenkapital zu hinterlegen. Dabei stellen die Bonitätseinstufung der Kreditnehmer durch eine externe Ratingagentur oder interne Ratings das wesentliche Kriterium dar, nach der die Eigenkapitalhinterlegung ermittelt wird. Damit sind auch Kreditinstitute und Bausparkassen gefordert, ihre Kreditnehmer zu bewerten. Allerdings können Bausparkassen nicht auf externe Ratings zurückgreifen, da die meist aus dem amerikanischen Raum stammenden Ratingagenturen nicht mit dem System des Bausparens vertraut und somit keine Bewertungen verfügbar sind.

Im Folgenden wird ein logisches Modell zur Klassifikation ausgefallener und endgetilgter Bauspardarlehen und zur Ermittlung von Kreditausfallwahrscheinlichkeiten vorgestellt. Da die Modellierung auf einem speziellen algebraischen Ansatz basiert, wird das Modell als verbandstheoretisches Implikationenmodell bezeichnet.

Bausparspezifische Besonderheiten

Im Bereich der Kreditwürdigkeitsprüfung von Privatkunden und Unternehmen existieren zahlreiche Untersuchungen mit unterschiedlichen Modellierungsansätzen. Bei Privatkundenkrediten stehen Merkmale wie zum Beispiel Einkommen, Beruf, Darlehenshöhe und Alter im Vordergrund, die bei Vergabe des Darlehens ausgewertet werden. Bei Bauspardarlehen kann diese stichtagsbezogene Bonitätsprüfung erweitert werden um Auswertungen der

Historie des Bausparvertrags wie Spardauer, Anteil der Wohnungsbauprämie am Guthaben oder die Tarifwahl.

Für die Untersuchung wurden zirka drei Millionen Konten einer realen Bausparkasse aus den Jahren 2000 bis 2004 verwendet. Davon fanden rund 350 000 Bausparverträge Berücksichtigung, die die Sparphase bereits verlassen und keine Vor- und Zwischenfinanzierung erhalten haben. Aus diesen Konten wurden die endgetilgten und die ausgefallenen Bauspardarlehen herausgefiltert.

Dabei gilt ein Darlehen als endgetilgt, wenn das Darlehen innerhalb der Regeltilgungszeit zurückgezahlt worden ist. Es gilt gemäß dem Baseler Ausschuss für Bankenaufsicht [Base04] als ausgefallen, wenn im Untersuchungszeitraum ein Zahlungsverzug von mehr als 90 Tagen auftritt. Zahlungsverzüge werden dabei auf Basis der von der Bausparkasse vertraglich festgelegten Mindesttilgung ermittelt, die abhängig vom jeweiligen Tarif ist.

Grundidee des Ansatzes ist es, ein logisches Klassifikationsmodell zu entwickeln, das in der Lage ist, die Datensätze anhand eines Regelwerks zu klassifizieren.

Die Autoren

Dr. Petra Fakler ist Business Analystin im Bereich Finance & Credit bei der HSBC Trinkaus & Burkhardt AG, Düsseldorf, und **Prof. Dr. Rainer Schrader** lehrt am Zentrum für Angewandte Informatik, Universität zu Köln.

Dabei soll das Modell folgenden Ansprüchen genügen:

- Überschaubarkeit der Regelsätze
- Transparenz für den Anwender
- Erfassung vieler Datensätze mit den Regeln
- Ermittlung von Wahrscheinlichkeiten für verschiedene Risikoklassen
- Übertragbarkeit auf andere Fragestellungen

Da die Datensätze umfangreiche Informationen enthalten, muss das Ziel sein, diejenigen Merkmale und Merkmalskombinationen zu extrahieren, die für die Klassenzugehörigkeit eines Datensatzes verantwortlich sind. Im Folgenden wird dazu die Theorie der formalen Begriffsanalyse zur Modellbildung verwendet. Diese mathematische Theorie (vergleiche [GWi96]) bietet eine Vielzahl von Ansätzen, um Zusammenhänge und Strukturen in Daten zu beschreiben, die sich vor allem durch ihre Transparenz für den Anwender auszeichnen.

Ermittlung der signifikanten Merkmalskombinationen

Um die zum Teil nicht binären Daten einer logischen Modellierung zugänglich zu machen, müssen sie diskretisiert werden. Dabei orientieren sich die Intervalle der Diskretisierung an der inhaltlichen Bedeutung der Merkmale. So wird beispielsweise die Spardauer mittels der von der Bausparkasse vorgeschlagenen Regel-sparrate diskretisiert.

Die so bearbeiteten Daten lassen sich in Form einer $(0,1)$ -Matrix darstellen. Die Zeilen dieser Matrix entsprechen den Bausparkonten und werden mit G (für Gegenstände) bezeichnet. Die Spalten entsprechen Merkmalen (daher als M bezeichnet), die die Konten besitzen oder nicht besitzen.

Anhand dieser Matrix, die wir im Folgenden auch als (formalen) Kontext bezeichnen, sollen für die Klassifikation relevante Merkmale beziehungsweise Merkmalskombinationen ermittelt werden. Die Klassifikation in endgetilgte beziehungsweise ausgefallene Verträge zerlegt die Zeilen in zwei Teilmengen G^+ und G^- und zugeordneten Kontexten K^+ und K^- . Es sollen Regeln formuliert werden, welche die strukturellen Zusammenhänge innerhalb beider Teilmatrizen

Abbildung 1: Auszüge aus dem Regelwerk zur Klassifikation endgetilgter und ausgefallener Bauspardarlehen

<p>Signifikante Regeln zur Erkennung ausgefallener Bauspardarlehen: Weiteres Konto in Zahlungsschwierigkeiten Selbstständig → Tarif 1/Tarif 3 Hohes Darlehen \wedge Selbstständig → Geringe WoP Arbeiter \wedge Geringes Alter → Tarif 1/Tarif 3 Hohe Spardauer \wedge Mittleres Alter → Geringe WoP</p> <p>Signifikante Regeln zur Erkennung endgetilgter Bauspardarlehen: Weiteres Konto nicht in Zahlungsschwierigkeiten Geringe Spardauer \wedge Hohes Alter → Kleines Darlehen Beamter → Geringe WoP Nicht Selbstständig \wedge Hohes Alter → Kein Arbeiter Rentner → Tarif 1/Tarif 3</p>

erfassen. Triviale Merkmale, das heißt Merkmale die keinen Beitrag zur Klassifikation leisten, sollen ebenfalls erkannt werden beziehungsweise im Regelwerk nicht mehr erscheinen.

Um Merkmalszusammenhänge zu erarbeiten, stellt die formale Begriffsanalyse verschiedene Methoden zur Verfügung. Bei einer hohen Anzahl von Gegenständen und einer relativ kleinen Menge von Merkmalen bietet es sich an, den Kontext anhand der Implikationen zwischen den Merkmalen zu erschließen [GaWi96]. Darunter verstehen wir Aussagen der Art: „Jeder Gegenstand mit den Merkmalen a,b,c, \dots hat auch die Merkmale x,y,z, \dots . Formal ist eine Implikation zwischen Merkmalen ein Paar von Teilmengen der Merkmalsmenge M . Wir bezeichnen eine solche Implikation mit $A, B, (A, B \in M)$. Die Implikationen geben Aufschluss über die Struktur des Kontextes.

Zu jedem Kontext mit endlicher Merkmalsmenge M gibt es eine vollständige und nichtredundante Menge von Implikationen (siehe [DuGu86]), die sogenann-

te Stammbasis, die sich algorithmisch ermitteln lässt (vergleiche unter anderem [Lind99]).

Ermittlung der Stammbasis

Da die Stammbasen der Kontexte K^- und K^+ noch zu umfangreich sind, werden in einem weiteren Schritt triviale Implikationen entfernt. Danach wird mit Hilfe eines X^2 -Unabhängigkeitstest überprüft, inwieweit die verbliebenen Implikationen der Kontexte K^+ und K^- signifikant für die jeweiligen Mengen K^+ und G^- sind. Daraus ergeben sich zwei Mengen von signifikanten Implikationen (im Weiteren als Regeln bezeichnet) mit zugehörigen X^2 -Werten. Diese Regeln müssen noch zu einem handlichen Regelwerk transformiert werden.

Im Hinblick auf die Anwendung des Modells im Rahmen von IRB-Ansätzen und die formulierten Grundanforderungen soll die Zusammenfassung weiterhin eine vorher festgelegte Anzahl von Teilmengen erzeugen, die anschließend mit dem

vorhandenen Datenmaterial empirisch bewertet werden können. Dies entspricht der Bildung unterschiedlicher Bonitätsklassen, wie sie von Basel II gefordert wird. Weiter soll das Verfahren einfach zu implementieren sein und zudem die Stärke der einzelnen Regeln berücksichtigen.

Die Regeln liegen nach ihrem X^2 -Wert geordnet vor. Um zu gewährleisten, dass eine vorab festgelegte Anzahl an Teilmengen entsteht, spalten wir die Implikationen anhand des Medians der X^2 -Werte auf. Dies liefert eine Halbierung in stark signifikante beziehungsweise signifikante Teilmengen jeweils für G^+ und G^- . Abbildung 1 zeigt einen Auszug aus dem so ermittelten Regelwerk. Dabei handelt es sich um einzelne Klauseln der zusammengefassten Regeln.

Die Zusammenfassung und Kombination der Regeln liefert eine natürliche Zerlegung des Bestandes in unterschiedliche Klassen. Zusätzlich kann bei Anwendung des Modells im Rahmen der Klassifikation gezielt jede vom Datensatz erfüllte Regel mit ihrem zugehörigen X^2 -Wert betrachtet werden. Ein Datensatz wird nun genau der maximalen Menge der erfüllten Regeln zugeordnet. Dadurch wird eine eindeutige Zuordnung der Datensätze zu den Teilmengen gewährleistet. Dieser Sachverhalt kann durch einen einfachen Entscheidungsbaum dargestellt werden. Da die a-priori-Wahrscheinlichkeiten für beide Klassen bekannt sind, kann für jede Teilmenge, beziehungsweise Bonitätsklasse mit Hilfe des Satzes von Bayes die bedingte Wahrscheinlichkeit für die Zugehörigkeit zur jeweiligen Klasse ermittelt werden. Damit erhalten wir Bonitätsklassen mit zugehörigen Ausfallwahrscheinlichkeiten.

Verwendung des Modells im Rahmen von IRB-Ansätzen

Der Baseler Ausschuss für Bankenaufsicht stellt in [Base04] einige Anforderungen an Ratingsysteme zur Ermittlung von Bonitätsklassen und Ausfallwahrscheinlichkeiten. Abbildung 2 zeigt beispielhaft einige wichtige Anforderungen und Kriterien an Ratingverfahren zur Verwendung in IRB-Ansätzen und deren Umsetzung im verbandstheoretischen Implikationenmodell. Die Angaben in den eckigen Klammern beziehen sich auf die zugehörigen Absätze in [Base04].

Die Ergebnisse des verbandstheoretischen Implikationenmodells wurden mit anderen Klassifikationsverfahren verglichen. In den Vergleich einbezogen wurden neuro-

Abbildung 2: Ausgewählte Mindestanforderungen und Kriterien an Ratingverfahren zur Verwendung in IRB-Ansätzen und deren Umsetzung im verbandstheoretischen Implikationenmodell

Mindestanforderung an ein Ratingverfahren	Umsetzung im verbandstheoretischen Implikationenmodell (VI)
Einteilung der Kreditnehmer in mindestens acht Bonitätsklassen [404]	Das VI erzeugt 15 Klassen von Kreditnehmern mit unterschiedlichen Merkmalsausprägungen
PD-Schätzung für jede Bonitätsklasse [409]	Empirische Ermittlung von bedingten Wahrscheinlichkeiten für jede Klasse
Gleichmäßige Verteilung der Kreditnehmer über die Ratingklassen [406]	Wird durch die Zusammenfassung der Implikationen erreicht
Detaillierte Beschreibung der Risikoklassen, konsistente Zuordnung der Risiken zu den jeweiligen Klassen [410]	Konsistente Zuordnung der Kreditnehmer durch Abfrage der logischen Regeln, welche durch die enthaltenen Implikationen detailliert beschrieben sind
Detaillierte, nachvollziehbare und interpretierbare Ratingdefinition [410]	Das logische Regelwerk im VI ist sehr gut interpretierbar
Kriterien müssen mit bankinternen Kreditvergabeberichtlinien übereinstimmen [410]	Die inhaltliche Diskretisierung der rationalen Merkmale erlaubt eine Einbeziehung interner Verfahren
Zeithorizont für die PD-Schätzung von mindestens einem Jahr [414]	Empirische Ermittlung der PD auf Grundlage von fünf Jahren

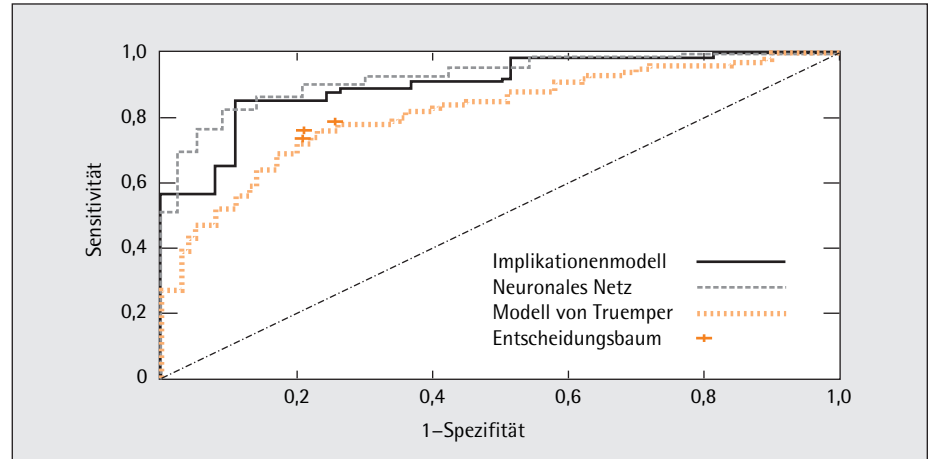
nale Netze, Entscheidungsbäume sowie das logische Modell von Truemper (vergleiche [Zim94, Quin93, True04]).

In einem ersten Schritt wurden dazu entsprechende Klassifikatoren anhand eines realen Darlehenskollektives (Ausgangsbausparkasse) entwickelt. Für die Ausgangsbausparkasse zeigt Abbildung 3 die Ergebnisse aller vier untersuchten Modelle in Form von ROC-Graphen (vergleiche [Fawc03, MaPr04]). Es zeigt sich, dass das neuronale Netz die höchste Klassifikationsgüte für die Ausgangsbausparkasse liefert, gefolgt vom verbandstheoretischen Implikationenmodell, den Entscheidungsbäumen und dem Modell von Truemper.

Zur Überprüfung der Generalisierungsfähigkeit werden alle Modelle ohne weitere Anpassungen auf die Originaldaten einer weiteren Bausparkasse (Validierungsbausparkasse) angewendet. Eine Übertragbarkeit des Modells wäre von Vorteil, da somit eine kosten- und zeitintensive Modellierung in den einzelnen Bausparkassen entfallen könnte. Abbildung 4 zeigt die Ergebnisse aller vier Modelle für die Validierungsbausparkasse. Dabei fallen die neuronalen Netze gerade durch ihre schlechte Klassifikationsgüte auf. Die höchsten Klassifikationsergebnisse liefert das verbandstheoretische Implikationenmodell, gefolgt von den Entscheidungsbäumen und dem Modell von Truemper. Das verbandstheoretische Implikationenmodell liefert somit robuste Klassifikationsergebnisse für die Ausgangs- und Validierungsbausparkasse.

Für die Kollektiventwicklung einer Bausparkasse ist die Darlehensnehmerquote von zentraler Bedeutung. Wir haben den verbandstheoretischen Ansatz verfolgt,

Abbildung 3: Vergleich der Klassifikationsergebnisse aller Modelle für die Ausgangsbausparkasse



um determinierende Faktoren für den Darlehensverzicht beziehungsweise die Darlehensnahme herzuleiten. Da die Darlehensannahme jedoch auch stark von exogenen Faktoren beeinflusst wird (unter anderem dem Außenzins, vergleiche etwa [Chev05]), haben wir hierzu ausschließlich Datensätze eines Jahres mit geringen Zinsschwankungen ausgewählt. Mit Hilfe des verbandstheoretischen Implikationsmodells konnten zusätzliche Faktoren identifiziert werden, die auf Darlehensverzichte beziehungsweise Darlehensannahme hindeuten.

Vorteile des Modells

Ein Vorteil des verbandstheoretischen Implikationsmodells liegt in der Erzeugung eines für den Anwender überschaubaren und nachvollziehbaren Regelwerks. Zudem liefert das Modell aufgrund der Zusammenfassung der Regeln auf natürliche Weise eine Zerlegung des Darlehenbestandes in unterschiedliche Boni-

tätsklassen. Die Anzahl der Bonitätsklassen kann dabei über den Grad der Zusammenfassung der Regeln gesteuert werden.

Durch Angabe der zugrunde liegenden Regeln sind die Ratingklassen klar definiert und ordnen Kreditnehmer mit gleichen Merkmalen derselben Bonitätsklasse zu. Auch die hohe Generalisierungsfähigkeit und die erfolgreiche Anwendung auf eine weitere bausparteknische Fragestellung sind von Vorteil. Zudem lieferte das verbandstheoretische Implikationenmodell für alle untersuchten Problemstellungen relativ robuste Ergebnisse.

Literaturverzeichnis

[Base04] Baseler Ausschuss für Bankenaufsicht: Internationale Konvergenz der Kapitalmessung und Eigenkapitalanforderung, Basel 2004.
 [Chev05] Chevalier, Thomas: Ein Risikomodell für Bausparkollektive. Dissertation, Köln 2005.
 [DuGu86] Duquenne, Vincent; Guigues, Jean-Louis: Familles minimales d'implications informatives d'un tableau de données binaires. In: Mathématiques et Sciences Humaines 24 (1986) 95, S. 5-18.
 [GaWi96] Ganter, Bernhard; Wille, Rudolf: Formale Begriffsanalyse: Mathematische Grundlagen. Springer Verlag, Berlin 1996.
 [Lind99] Lindig, Christian: Algorithmen zur Begriffsanalyse und ihre Anwendung bei Softwarebibliotheken. Dissertation, Braunschweig 1999.
 [True04] Truemper, Klaus: Design of Logic-Based Intelligent Systems. John Wiley & Sons, Inc., New Jersey, 2004.
 [Quin93] Quinlan, J. Ross: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo 1993.
 [Zim94] Zimmermann, Hans-G.: Neuronale Netze als Entscheidungskalkül. In: Rehkugler, Heinz, Zimmermann, Hans-G. (Hrsg.): Neuronale Netze in der Ökonomie. Vahlen, München 1994, S. 1-87.
 [Fawc03] Fawcett, Tom: ROC Graphs: Notes and Practical Considerations for Researchers. Technical report HP Laboratories, Palo Alto 2003.
 [MaPr04] Macskassy, Sofos, A.; Provost, Forster J.: Confidence Bands for ROC Curves: Methods and an Empirical Study. In: Hernandez-Orallo, Jose; Ferri, Cesar; Lachiche, Nicolas; Flach, Peter A. (Hrsg.): ROC Analysis in Artificial Intelligence. Proceedings, Valencia 2004, S. 61-70.

Abbildung 4: Vergleich der Klassifikationsergebnisse aller Modelle für die Validierungsbausparkasse

