

Anwendung der Dempster-Shafer Evidenztheorie auf die Bonitätsprüfung

Bertrand Henkel

Dr. Markus Höchstötter

Karlsruher Institut für Technologie

1 Einleitung

Geldinstitute müssen sich bei einer Kreditvergabe regelmäßig mit der Frage auseinandersetzen, wann einem potenziellen Kreditnehmer ein Kredit gewährt werden sollte.

Ein wichtiger Aspekt stellt hierbei die Bonitätsprüfung dar. Sie bewertet die Kreditwürdigkeit, indem sie die Ausfallwahrscheinlichkeit eines potenziellen Schuldners bestimmt. Die Bestimmung der Ausfallwahrscheinlichkeit eines Kreditnehmers ist essentiell für die Bonitätsprüfung.

Dempster und Shafer entwickelten eine (Evidenz-)Theorie, in der es möglich ist eine generelle Unsicherheit über die gegebenen Informationen anzunehmen. Sie stellt eine echte Erweiterung der Bayes-Theorie dar. Ziel dieser Arbeit ist es, unter Anwendung der Dempster-Shafer Evidenztheorie, eine weitere Möglichkeit neben der klassischen Bonitätsprüfung darzulegen.

Hierzu werden die Elemente *frame of discernment*, *basic probability mass* sowie die *Belief* und *Plausibilitäts-Funktion* der Dempster-Shafer Evidenztheorie eingeführt und anschließend auf die Bonitätsprüfung angewendet.

Anders als bei klassischen Verfahren gibt es bei dieser Theorie explizit ein *basic probability mass* für ein allgemeines Nichtwissen. Dies ist das wesentliche Unterscheidungsmerkmal bei Dempster und Shafer und hat weitreichende Konsequenzen. Eine dieser Konsequenzen ist, dass anstelle einzelner Wahrscheinlichkeiten nunmehr Wahrscheinlichkeitsintervalle auftreten.

2 Vorgehen und verwandte Arbeiten

Basierend auf den wissenschaftlichen Arbeiten von Ballal *Dempster Shafer Theory* [1], Parsons *Some qualitative approaches to applying the Dempster-Shafer theory* [12], Reineking *Belief Functions: Theory and Algorithms* [13], Smets *Belief functions* [15] und Zadeh *Fuzzy sets. Information and Control* [19] und *Review of a mathematical theory of evidence* [21], die alle die Theorie von Dempster und Shafer aufgreifen, werden in diesem Artikel zunächst die generellen Elemente dieser Theorie definiert. Anschließend wird mit Blick auf die Vorgehensweise des im Buch „Consumer Credit Models“ von Lyn C. Thomas beschriebenen Scoringverfahrens die Evidenztheorie von Dempster und Shafer auf die Bonitätsprüfung angewendet. Anstatt Odds und den daraus resultierenden Score zu definieren, werden stattdessen Gewichtsfunktionen über die einzelnen Merkmale des Kreditnehmers eingeführt, sogenannte *weight of evidence*, die mit dem Score gleichzusetzen sind. Eine Transferfunktion, wie in dem Buch „A mathematical theory of evidence“ von Shafer beschrieben, bildet die *weight of evidence* schließlich auf ein Intervall $[0; 1]$ ab. Das Maß für die Unsicherheit wird mithilfe des totalen Differentials der Gewichtsfunktionen berechnet.

3 Elemente der Evidenztheorie von Dempster-Shafer

Die Dempster-Shafer Evidenztheorie, auch Dempster-Shafer-Theorie genannt (im weiteren Verlauf wird Dempster-Shafer mit **DS** abgekürzt), lässt sich als eine Verallgemeinerung der Wahrscheinlichkeitstheorie auffassen.

3.1 Frame of Discernment, Basic Probability Mass und Belief - bzw. Plausibilitäts-Funktionen

Analog zum Wahrscheinlichkeitsmaß der klassischen Wahrscheinlichkeitstheorie wird das *frame of discernment* Ω , welches eine Menge von Alternativen oder Aussagen $\Omega = (A_1, \dots, A_n)$ bezeichnet, definiert. Eine Hypothese ist hierbei eine Untermenge $A \subseteq \Omega$ des Wahrnehmungsrahmens. Eine *Evidenz*, oder auch *Basismaß*, in der englischen Literatur auch *basic probability mass* (im nachfolgenden wird dafür die Abkürzung **BPM** benutzt), ist eine Funktion, die jeder Teilmenge X von Ω eine Gewichtung bzw. ein Vertrauen zuweist. Sei Ω ein Wahrnehmungsrahmen, so ist die Evidenz m durch

folgende Abbildung definiert:

$$m : 2^\Omega \rightarrow [0; 1],$$

wobei 2^Ω die Menge aller Teilmengen von Ω bezeichnet.

Darüber hinaus gilt:

1. $m(\emptyset) = 0$, das heißt der leeren Menge wird keine Evidenzmasse zugeordnet,
2. $\sum_{X \subseteq \Omega} m(X) = 1$, das heißt die Summe aller Evidenzen für einen Wahrnehmungsrahmen Ω summiert sich zu 1.

Neben den BPM werden zusätzlich Belief- und Plausibilitäts-Funktionen eingeführt. Eine *Belief-Funktion* wird dabei wie folgt definiert: Sei Ω ein Wahrscheinlichkeitsrahmen und m ein darauf definiertes Basismaß. Die Belief-Funktion ist eine Funktion $Bel : 2^\Omega \rightarrow [0; 1]$ mit

$$Bel(X) := \sum_{Y \subseteq X} m(Y).$$

Folglich ergibt sich eine Belief-Funktion $Bel(X)$ durch die Addition der Basismaße m , sofern die Hypothese eines jeden Basismaßes eine echte Teilmenge von X ist. Die Belief-Funktion stellt somit ein Mindestmaß für die Wahrscheinlichkeit der Hypothese X dar. Daher definiert die *Belief-Funktion* die untere Grenze einer unbekanntem Wahrscheinlichkeitsfunktion P für eine Teilmenge $X \subseteq \Omega$:

$$Bel(X) \leq P(X), \quad \forall X \subseteq \Omega.$$

In diesem Fall werden die Belief-Funktion $Bel(X)$ und Wahrscheinlichkeitsfunktion $P(X)$ als *kompatibel*¹ bezeichnet.

Die Belief-Funktion stellt somit die Wahrscheinlichkeitsmasse an Glauben dar, die mindestens einer Hypothese X zugeordnet werden kann. Im Gegensatz zur Bayes-Wahrscheinlichkeitstheorie muss die Additivität² der Belief-Funktionswerte jedoch nicht gegeben sein. Das bedeutet, dass der Glaube in eine Hypothese X und der Glaube in seine Gegenhypothese $\neg X$ kleiner als 1 sein kann:

¹aus dem Englischen „compatible“

²Additivität von Bayes: Sei P eine nach der klassischen Wahrscheinlichkeitstheorie definierte Wahrscheinlichkeitsfunktion einer Teilmenge A so gilt $P(A) + P(\neg A) = 1$.

$$Bel(X) + Bel(\neg X) \leq 1 \quad \forall X \subseteq \Omega.$$

Eine *Plausibilitäts-Funktion* wird hingegen wie folgt definiert: Sei Ω ein Wahrscheinlichkeitsrahmen und m ein darauf definiertes Basismaß. Die Plausibilitäts-Funktion ist eine Funktion $Pl : 2^\Omega \rightarrow [0; 1]$ mit

$$Pl(X) := \sum_{Y \cap X \neq \emptyset} m(Y).$$

Nach dieser Definition ergibt sich die Plausibilitäts-Funktion $Pl(X)$ durch die Addition der Basismaße m einer Hypothese Y , sofern die Schnittmenge der Mengen X und Y ungleich der leeren Menge ist. Sie ist dementsprechend die obere Grenze einer unbekanntenen Wahrscheinlichkeitsfunktion P für die Teilmenge $X \subseteq \Omega$:

$$Pl(X) \geq P(X), \quad \forall X \subseteq \Omega.$$

Sie zeigt demnach auf, wie wahrscheinlich eine Hypothese X maximal werden kann. Die Plausibilitäts-Funktion ist die Masse an Glauben, die nicht direkt dem Komplement der Hypothese X zugeordnet werden kann. Es beschreibt deshalb, wie plausibel diese Hypothese X ist.

Auch bei Plausibilitäts-Funktionen muss die Additivität von Bayes nicht gegeben sein, denn es gilt:

$$Pl(X) + Pl(\neg X) \geq 1 \quad \forall X \subseteq \Omega.$$

Der wesentliche Unterschied zwischen der Plausibilitäts-Funktion $Pl(A)$ und der Belief-Funktion $Bel(A)$ ist, dass $Pl(A)$ die komplette Glaubensmasse, welche der Menge A zugesprochen werden kann, misst. $Bel(A)$ hingegen misst die komplette Glaubensmasse, die auf die Menge A begrenzt ist.³ Sie misst demnach die Glaubensmasse, die auf jeden Fall der Menge A zugesprochen werden kann.

Sind nun eine Wahrscheinlichkeitsfunktion P für eine Hypothese $X \subseteq \Omega$ und die dazu kompatible Belief-Funktion $Bel(X)$ bzw. Plausibilitäts-Funktion $Pl(X)$ gegeben, so lässt sich folgendes Intervall für die Wahrscheinlichkeit $P(X)$ bestimmen:

$$P(X) \in [Bel(X); Pl(X)].$$

³Ballal (2004), [1]

Damit kann man mithilfe von DS bei jeder Schätzung einer Wahrscheinlichkeit ein Fehlerintervall dieser Schätzung mit angeben.

Eine eindeutige Wahrscheinlichkeitsfunktion P existiert nach der DS-Theorie genau dann, wenn für alle $X \subseteq \Omega$ gilt, dass die Belief-Funktion $Bel(X)$ mit der Plausibilitäts-Funktion $Pl(X)$ übereinstimmt. Gilt demnach

$$\sum_{Y \subseteq X} m(Y) = Bel(X) = Pl(X) = \sum_{Y \cap X \neq \emptyset} m(Y),$$

erhält man folglich

$$Bel(X) = Pl(X) \triangleq P(X).$$

3.2 Unsicherheitsintervalle

Eine Unsicherheit spiegelt sich in der Definition der Belief- bzw. Plausibilitäts-Funktionen als untere bzw. obere Grenze für eine kompatible Wahrscheinlichkeitsfunktion P wider. Diese beiden Funktionen definieren ein Wahrscheinlichkeitsintervall für die Teilmenge X . Man kann diese Grenzen als Unsicherheit über das Eintreffen eines Ereignisses aus der Teilmenge X auffassen. Formal ist die Unsicherheit $U(X)$ für eine Hypothese $X \subseteq \Omega$ wie folgt definiert:

$$U(X) := Pl(X) - Bel(X). \quad (1)$$

Durch Umstellung ergibt sich damit

$$Pl(X) = Bel(X) + U(X). \quad (2)$$

Die Unsicherheit entsteht sobald man keine Evidenz über ein Ereignis hat, die Quelle der Evidenz für ein Ereignis unsicher ist oder unvorhergesehene Zukunftereignisse, die die Evidenzen beeinflussen, möglich sind.

3.3 Vergleich

Durch die Definition der Plausibilitäts-Funktion $Pl(A) = Bel(A) + U$ kann man durch nachfolgende Grenzwertbildung zeigen, dass $Bel(A) = Pl(A) = P(A)$ gilt. Geht die Unsicherheit U nun gegen null, so existieren die Wahrscheinlichkeiten und Wahrscheinlichkeitsfunktionen nach der klassischen Bayes-Wahrscheinlichkeitstheorie, da das Wahrscheinlichkeitsintervall $P(A)$ bei Grenzwertbetrachtung immer kleiner wird und somit die Evidenz ohne jegliche Unsicherheit zur Wahrscheinlichkeit wird. Folglich gilt:

$$Pl(A) = Bel(A) + U \Leftrightarrow \lim_{U \rightarrow 0} Pl(A) = Bel(A).$$

Daraus folgt, dass die DS-Theorie einer Verallgemeinerung der Bayes-Theorie entspricht.

4 Bonitätsprüfung im Rahmen der Dempster-Shafer Evidenztheorie

Mit Anlehnung an das Buch „Consumer Credit Models“ von Lyn C. Thomas, welches unter anderem die klassische Vorgehensweise bei einer Bonitätsprüfung darlegt, wird im Nachfolgenden die DS-Theorie auf diese Prüfung angewendet.

Frame of Discernment

Im Rahmen der DS-Theorie muss zunächst der Wahrnehmungsrahmen definiert werden. Analog zum Buch „Consumer Credit Models“ werden die zwei Hypothesen des *frame of discernment* Ω wie folgt definiert:

- $G(\text{ood}) = \text{Hypothese, dass der KN ein guter KN sein wird,}$
- $B(\text{ad}) = \text{Hypothese, dass der KN ein schlechter KN sein wird.}$

Im Unterschied zur Bayes-Wahrscheinlichkeitstheorie wird zusätzlich eine neue Hypothese $G \vee B$ (die Oder-Kombination "Good or Bad") eingeführt:

- $G \vee B = \text{Hypothese, dass der KN entweder ein guter oder ein schlechter KN ist. Für } G \vee B \text{ wird analog die Bezeichnung } \mathbf{U} \text{ (Unsicherheit) benutzt.}$

Jede Evidenz, die in diese dritte Hypothese einfließt, beschreibt die allgemein vorliegende Unsicherheit. Es gilt nun:

$$m(G) + m(B) + m(G \vee B) = m(G) + m(B) + m(U) = 1. \quad (3)$$

Definition der Basiswahrscheinlichkeitsmaße $m(G)$ und $m(B)$

Es soll zunächst das BPM $m(G)$ betrachtet werden. Analoges gilt jedoch auch für das BPM $m(B)$. Der Fall des BPMs $m(U)$, das Maß für die Unsicherheit, wird gesondert betrachtet.

Das BPM für die Hypothese „Good“ $m(G)$ hängt über $\lim_{m(U) \rightarrow 0} m(G) = p(G | X)$ mit der bedingten Wahrscheinlichkeit zusammen.⁴

Der Merkmalsvektor $X = (x_1, x_2, \dots, x_n)$, wobei $n \in \mathbb{N}$ die Anzahl der Ausprägungen ist, stellt dabei die einzelnen Merkmale x_i , mit dem Kreditnehmer bewertet oder klassifiziert werden, dar (mit $i \in \mathbb{N}$ mit $i \leq n$ gilt).⁵ Dies bedeutet, dass der Vektor X die einzelnen Merkmalsausprägungen des KNs beschreibt. Die einzelnen Merkmale müssen dabei jeweils dimensionslos sein. Dies wird dadurch erreicht, dass man entweder das ursprüngliche Merkmal durch einen Referenzwert des Merkmals teilt oder die einzelnen Merkmalsausprägungen durch entsprechende Zahlenwerte repräsentiert. Mithilfe von

$$x_k = \frac{f_k}{\max_k |f_k|}$$

werden die einzelnen Merkmale schließlich standardisiert, f_k bezeichnet dabei das ursprüngliche entdimensionalisierte k -te Merkmal.

Damit gilt $x_k \leq |1|$. Sie werden so begriffen, dass sie zu Evidenzen im Sinne von DS führen.

Definition der Gewichtsfunktion w_G

Aus der Praxis kann entnommen werden, dass nicht alle Merkmale gleich stark in die Bewertung eines KNs in „Good“ oder „Bad“ mit einfließen.⁶

Das motiviert zu jedem Merkmal k einen Gewichtungsfaktor g_k zu definieren, sodass die einzelnen Merkmale die Gewichtsfunktion bzw. den Score unterschiedlich, je nach Gewichtung g_k , beeinflussen.

Aus diesem Grund wird zunächst jeweils eine Gewichtsfunktion w_G bzw. w_B für die Hypothesen G und B bestimmt und anschließend die Transferfunktion $T(w(X)) = 1 - e^{(-w(X))}$ jeweils darauf angewendet, um den Wertebereich $[0; \infty]$ der Gewichtsfunktion auf den Bereich $[0; 1]$ abzubilden. Diese Transferfunktion entstammt dem Buch von Shafer⁷ und kann mit dem BPM $m(A)$

⁴Betrachte die Vorgehensweise mit den bedingten Wahrscheinlichkeiten $p(G | X)$ im Buch „Consumer Credit Models“ [17].

⁵So können beispielsweise sozialwirtschaftliche Eigenschaften wie Alter, Wohnstatus oder Arbeitsverhältnis, sowie die vergangene Kreditperformance, die Informationen von verspäteten oder verpassten Zahlungen beinhaltet, und die existierenden Schuldverpflichtungen als Merkmale betrachtet werden (aus dem Buch „Consumer Credit Models“ [17]).

⁶Bundesverband deutscher Banken (2006), [3], Seite 9

⁷Shafer (1976), [14], Seiten 77 -78, *Weight of Evidence*

zur Hypothese A gleichgesetzt werden.

Mithin gilt

$$m(G) = T(w(X)) = 1 - e^{(-w(X))}.$$

Hierbei wird der Exponent der E-Funktion als *weight of evidence* (nachfolgend mit **WOE** abgekürzt) definiert. Daher kann man diese Gewichtsfunktion w_G als WOE bezeichnen und mit dem klassischen Score (aus dem Buch) gleichsetzen. Die Definition der Gewichtsfunktion w_G durch den Merkmalsvektor X sieht wie folgt aus:

$$w_G = \sum_{k=1}^n g_k \cdot x_k. \quad (4)$$

Die einzelnen Summanden dieser Gewichtsfunktion w_i können als die WOE_i für das i -te Merkmal aufgegriffen werden. Die einzelnen g_k sind die Gewichtungen des jeweiligen k -ten Merkmals für die Gewichtsfunktion, weshalb $g_k \in [0; 1]$ gilt. Sie bestimmen daher den Einfluss des Merkmals x_k auf die Gewichtsfunktion w_G .

Wählt man nun eines der Merkmale x_s als primär oder wichtigstes Merkmal aus und setzt dessen Gewichtungsfaktor g_s gleich 1, so kann man auch alle anderen Merkmale $i \neq s$ über den Gewichtungsfaktor g_i als äquivalentes Primärmerkmal x_s auffassen.

Die WOE kann man dann als Summe der Merkmale in Einheiten des Primärmerkmals berechnen. Dies hat den Vorteil, dass man alle Merkmale beispielsweise in die Einheit „Einkommen“ umrechnen kann. Mithin lassen sich nun die einzelnen WOE besser vergleichen. Zudem lässt sich nun bestimmen, wie viel zusätzliches Primärmerkmal man benötigt, um einen bestimmten Wert für die WOE zu erhalten.

Durch die Transferfunktion werden die Gewichtsfunktionen w_G und w_B auf das Intervall der BPM $[0; 1]$ abgebildet.

Dabei ist die Normierung der BPM für ein gegebenes WOE $W_{\frac{1}{2}}$ festzulegen.⁸ Dies geschieht durch einen allgemeinen Skalierungsfaktor g_0 , der festlegt, dass ein $W_{\frac{1}{2}}$ auf ein BPM $m = 0.5$ abgebildet wird.

Es soll also gelten:

$$m(G) = 1 - e^{-g_0 \cdot W_{\frac{1}{2}}} = 0.5.$$

⁸ $W_{\frac{1}{2}}$ ist dabei ein bestimmter fester Wert der Gewichtsfunktion w_G .

g_0 ergibt sich dann durch Umformungen zu:

$$g_0 = \frac{-\ln(0.5)}{W_{\frac{1}{2}}} = \frac{\ln(2)}{W_{\frac{1}{2}}}.$$

$W_{\frac{1}{2}}$ könnte man zunächst willkürlich oder empirisch festlegen. Versteht man aber $W_{\frac{1}{2}}$ in Einheiten des Primärmerkmals, so legt man durch g_0 fest, wie viel Einheiten des Primärmerkmals notwendig sind, um ein BPM von $m(G) = 0.5$ zu erreichen.

Ergänzend soll erwähnt werden, dass immer die Normierungsbedingung

$$m(G) + m(B) + m(U) = 1 \tag{5}$$

gilt.

Um ein vom BPM $m(G)$ verschiedenes BPM $m(B)$ zu erhalten, sollten die Gewichtungen g_k , die in Gleichung (4) definiert werden, verändert werden. Der Grund liegt darin, dass ein schlechter KN andere Gewichtungen der einzelnen Merkmale besitzen kann, da man auf andere Merkmale größeren Wert legt. Es wird jedoch in der Schreibweise der Gewichtungen g_k zunächst keine Unterscheidung zwischen einem guten und einem schlechten KN gemacht.

Es ist jedoch zu beachten, dass die Summe der beiden ersten BPM $m(G)$ und $m(B)$ nach ihrer Definition bereits größer 1 sein könnte. Deshalb sind die oben definierten BPM's als vorläufige, nicht normierte Funktionen zu betrachten, die schließlich wie folgt normiert werden (Renomierungsschritt):

$$Z = m^{un}(G) + m^{un}(B) + m^{un}(U),^9$$

$$m(G) = \frac{m^{un}(G)}{Z}, m(B) = \frac{m^{un}(B)}{Z}, m(U) = \frac{m^{un}(U)}{Z}.$$

Die Unsicherheit $m(U)$ spiegelt in diesem Fall die generell vorherrschende Unsicherheit über die Güte des KNs wider. Diese generelle Unsicherheit ergibt sich beispielsweise aus folgenden Gründen:

- Fehlen von Daten,
- unsichere Datenlage,
- keine eindeutigen Daten,

⁹wobei $m^{un}(A)$ für das nicht normierte BPM der Hypothese A steht (*un = nicht normiert*).

- unvorhergesehene Zukunftsereignissen.¹⁰

Bestimmung der Unsicherheit $m(U)$

Die Unsicherheit wird mit dem Ansatz des totalen Differentials bestimmt und auf die Unsicherheiten des Merkmalsvektors zurückgeführt. Es gilt

$$m(U) = \sqrt{dm(G, X)^2 + dm(B, X)^2} + \Delta m_U. \quad {}^{11}$$

Im Folgenden wird lediglich $dm(G, X)$ berechnet, da die Berechnung von $dm(B, X)$ analog erfolgt.¹²

Die Definition von $m(G) = 1 - e^{-g_0 \cdot w_G}$ mit der WOE für „Good“, gegeben dem Merkmalsvektor X , $w_G = \sum_{k=1}^n g_k \cdot x_k$ ist bereits bekannt. Für g_0 gilt weiterhin $g_0 = \frac{\ln(2)}{W_{\frac{1}{2}}}$.

Daher kann nun das totale Differential von $m(G)$ gebildet werden, indem man diese Funktion jeweils nach x_i ableitet. Es wird demnach partiell nach jedem Attribut x_i des Merkmalsvektors $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ abgeleitet, sodass man

$$dm(G, X) = e^{-g_0 \cdot w_G} \cdot (-g_0) dw_G$$

erhält, wobei $dw_G = \sum g_k \cdot dx_k$ das totale Differential von w_G beschreibt. dx_k bezeichnet die Abweichung von dem Merkmal x_k . Zudem kann dw_G als die Unsicherheit der WOE interpretiert werden, die letztlich durch die Unsicherheit der Merkmale und die Unsicherheit bei der Bestimmung der Gewichtungen g_k gegeben ist.

Mit der Umformung $-e^{-g_0 \cdot w_G} = m(G) - 1$ und dem Setzen von $F = (g_0) \cdot dw_G$ erhält man nun für das totale Differential $dm(G, X) = (1 - m(G)) \cdot F$.

F ist ein fester Wert, der jedoch abhängig von dem Merkmalsvektor X ist. Mit analogen Rechnungen für die Hypothese B(ad), erhält man $dm(B, X) = (1 - m(B)) \cdot F$.

¹⁰Zu erwähnen sind beispielsweise ein Autounfall, ein Rohrbruch im Haus, ein Lottogewinn oder eine überraschende Bonusausschüttung als zusätzliches Gehalt.

¹¹Damit wird sichergestellt, dass die Unsicherheit stets positiv bleibt. Der Ausdruck $dm(G, X)$ wird nachfolgend definiert und beschreibt die Ableitung von $m(G)$ nach den Merkmalsvektor X .

¹²Zu beachten ist, dass zunächst angenommen wird, dass die Unsicherheit der WOE dw bei G und B gleich ist. Dies ist jedoch bei unterschiedlichen Gewichtungsfaktoren g_k nicht der Fall. In diesem Fall müsste mit einem $F = F_G + F_B$ gerechnet werden, wobei $F_G = g_0 \cdot dw_G$ bzw. $F_B = g_0 \cdot dw_B$ und dw_G bzw. dw_B für das totale Differential für w_G bzw. w_B steht.

Durch Einsetzen in die obige Definition der Unsicherheit $m(U)$ erhält man schließlich

$$\begin{aligned} m(U) &= \sqrt{dm(G, X)^2 + dm(B, X)^2} + \Delta m_U \\ &= \sqrt{((1 - m(G)) \cdot F)^2 + ((1 - m(B)) \cdot F)^2} + \Delta m_U. \end{aligned}$$

Diese Unsicherheit basiert auf der Überlegung, dass einerseits die Merkmale für den KN geschätzt werden müssen und andererseits dass eine generelle Unsicherheit der Merkmale vorherrscht. Im Laufe der Kredittilgungsphase können sich Merkmale beispielsweise durch die wirtschaftliche Lage verändern. Mithilfe des Differentials lässt sich genau diese Abweichung vom ursprünglichen Merkmalswert darstellen. Denn das Differential $dm(W, x)$ zeigt die Veränderung der Funktion $W(x)$ an, die man erhalten würde, wenn man ein kleines Δx in x -Richtung fortschreitet. Dieses Δx ist mit den kleinen möglichen Veränderungen bzw. Unsicherheiten des Merkmalsvektors gleichzusetzen. Dadurch erhält man die Unsicherheit $m(U)$ anhand der Unsicherheit des Merkmalsvektors X , die sich letztlich in den Basiswahrscheinlichkeiten $m(G)$ bzw. $m(B)$ widerspiegelt.

Der Wert $\Delta m_U \geq 0$ spiegelt eine generelle Unsicherheit über die Güte des KNs bzw. eine Fehlerfunktion wider. Er stellt zudem eine unvorhergesehene Veränderung in den einzelnen Merkmalen dar. Durch die Einführung dieser Konstante wird ebenfalls verhindert, dass die Unsicherheit $m(U)$ gegen null konvergiert. Die Konstante Δm_U könnte man beispielsweise durch empirische Daten bestimmen. Mit der Bestimmung der Unsicherheit m_U ist die Definition eines Basiswahrscheinlichkeitsmaßes vollständig, und es gilt die Normierungsgleichung (Gleichung (5)).

Kombination zweier Basiswahrscheinlichkeitsmaße

Die DS-Theorie stellt des Weiteren einige Möglichkeiten dar, mit denen man unterschiedliche Basismaße kombinieren kann. Darunter sind unter anderem *Dempster's Kombinationsregel* oder *Yager's Kombinationsregel* zu nennen. Der Unterschied zwischen den beiden Regeln ist, dass *Yager's Kombinationsregel* ein Basismaß außerhalb des *frame of discernment* annimmt, das heißt der leeren Menge ein Basismaß größer null erlaubt. Diese Regeln und die bestehende Problematik wird ausführlich unter anderem in den wissenschaftlichen Arbeiten von Orponen [11], Smets [15]; [16], Parsons [12] und Reineking [13] behandelt.

4.1 Beispiel der Kreditvergabe mithilfe der DS-Theorie

Bezugnehmend auf die Vorgehensweise im Buch „Consumer Credit Models“ von Lyn C. Thomas wird nun die DS-Theorie auf die Bonitätsprüfung angewendet. Dabei wird das sogenannte Expected Monetary Value -Kriterium, kurz EMV-Kriterium, nach dem eine Bank einem potenziellen KN einen Kredit gewähren sollte, sobald der Profit der Bank durch das Akzeptieren des Kredits größer als null ist, verwendet:

$$f \cdot [g \cdot p_G - l \cdot (1 - p_G)] - (1 - f) \cdot c > 0.^{13}$$

Diese Formel lässt sich leicht durch verschiedene Umformungen in eine Schwelle S_{Bayes} ¹⁴ für die Wahrscheinlichkeit, dass ein KN ein guter ist, überführen

$$p_G > \left(\frac{1}{f} - 1\right) \cdot \frac{c}{g+l} + \frac{l}{g+l} =: S_{Bayes}.$$

Diese Schwelle lässt sich folgendermaßen interpretieren: Sobald die Wahrscheinlichkeit p_G eines potenziellen KNs höher als die Schwelle S_{Bayes} ist, sollte die Bank den Kredit gewähren, da sie dadurch einen Profit größer null erreichen kann.

Gemäß der DS-Theorie liegt die Wahrscheinlichkeitsfunktion p_G im Intervall $p_G \in [Bel_G; Pl_G] = [Bel_G; Bel_G + U]$. Man kann nun einen Skalierungsfaktor $i \in [0; 1]$ einführen, und damit p_G wie folgt definieren

$$p_G^i := (1 - i) \cdot Bel_G + i \cdot Pl_G = Bel_G + i \cdot U.$$

Analog kann man die Wahrscheinlichkeit p_B mit der Belief-Funktion Bel_B , der Unsicherheit U und dem Skalierungsfaktor $j \in [0; 1]$ definieren: $p_B^j := (1 - j) \cdot Bel_B + j \cdot Pl_B = Bel_B + j \cdot U$. Aufgrund der Normierungsbedingung (Gleichung (5)) und wegen $m(G) = Bel_G$ bzw. $m(B) = Bel_B$ gilt $i + j \leq 1$.

¹³Die Variablen sind dabei wie folgt definiert:

- $p_G \hat{=}$ Wahrscheinlichkeit, dass der KN ein *guter* ist.
- $g \hat{=}$ Gewinn, den die Bank durch Gewährung des Kredits realisieren kann. Es gilt $g > 0$,
- $l \hat{=}$ Verlust, den die Bank durch Ausfall des KNs machen kann. Es gilt $l > 0$,
- $f \hat{=}$ Wahrscheinlichkeit mit der ein potenzieller KN das Angebot einer Bank für einen Kredit annimmt.
- $c \hat{=}$ Verwaltungsgebühren bei Ablehnung des Angebots durch den KN.

¹⁴ S_{Bayes} steht für die Schwelle nach der klassischen Wahrscheinlichkeitslehre nach Bayes.

Das EMV-Kriterium kann nun durch Einsetzen und Umformungen wie folgt umgeschrieben werden:

$$g \cdot Bel_G - l \cdot Bel_B > Konst - U \cdot (i \cdot g - j \cdot l),$$

wobei $Konst := (\frac{1}{f} - 1) \cdot c$ definiert. Außerdem gilt nach der DS-Theorie $Bel_B = 1 - Pl_G = 1 - Bel_G - U$. Somit lässt sich dieses Kriterium nur mithilfe der Belief-Funktion Bel_G und der Unsicherheit U ausdrücken:

$$Bel_G > S_{Bayes} - K_{DS}(i, j), \quad (6)$$

wobei $S_{Bayes} = \frac{Konst}{g+l} + \frac{l}{g+l} = (\frac{1}{f} - 1) \cdot \frac{c}{g+l} + \frac{l}{g+l}$ die Schwelle nach der Bayes-Wahrscheinlichkeitstheorie und

$K_{DS}(i, j) := \frac{U \cdot [i \cdot g + (1 - j) \cdot l]}{g+l}$ die Korrektur zur Bayes-Schwelle gemäß DS bezeichnet. Die DS-Korrektur ist immer größer oder gleich null und hängt linear von den Skalierungsfaktoren i und j sowie von der Unsicherheit U ab. Ist $U = 0$, das heißt, herrscht keine Unsicherheit, erhält man die Bayes-Schwelle. Die Bayes-Schwelle resultiert auch, wenn $j = 1$ ist und damit $i = 0$, weil dann das Unsicherheitsmaß komplett der Wahrscheinlichkeit p_B zugeschlagen wird. Dies entspricht der klassischen Vorgehensweise der Kreditinstitute.

Wenn $i > 0$ ist, dann wird ein Teil der Unsicherheit als mögliche Wahrscheinlichkeit p_G akzeptiert, und die Akzeptanzschwelle wird proportional zu U gesenkt. Dies beschreibt das zusätzliche Gewinnpotenzial eines risikobereiteren Investors.

Je größer j ist, desto mehr Unsicherheit wird dem Wahrscheinlichkeitsmaß p_B zugesprochen, und die Schwelle erhöht sich proportional zu U .

Es ist auch möglich, dass $i + j < 1$ ist, sodass eine restliche Unsicherheit im System verbleibt. Der sich dadurch ergebende Korrekturfaktor ist stets größer null, sodass die Akzeptanzschwelle immer kleiner als nach der klassischen Vorgehensweise ist. Sie ist auch linear abhängig von der Unsicherheit U sowie dem Faktor $1 - i - j$.

Betrachtet man jedoch den Fall, dass die gesamte Unsicherheit U aufgeteilt wird, gilt $i + j = 1$, also $j = 1 - i$. Damit ändert sich die DS-Korrektur wie folgt:

$$K_{DS} = \frac{U \cdot [i \cdot g + (1 - (1 - i)) \cdot l]}{g+l} = i \cdot U.$$

Obige Ungleichung (6) lässt sich nun folgendermaßen interpretieren: Sobald ein potenzieller KN eine größere Belief-Funktion Bel_G als die Bayes-Schwelle abzüglich der DS-Korrektur hat, sollte die Bank den Kredit gewähren, da sie nun bei einem Darlehensgeschäft einen Profit erreichen kann. Der Korrekturfaktor K_{DS} beschleunigt in gewissem Maße die Akzeptanz eines potenziellen KNs. Umgekehrt geht der Kreditgeber ein höheres Verlustrisiko ein. Dieses wird er sich normalerweise durch eine höhere Gewinnchance, zum Beispiel ein höheres g für das Unsicherheitsmaß U , vergüten lassen.

Die Bedeutung der Skalierungsfaktoren i und j

Der Skalierungsfaktor i ist dabei mit der Erhöhung der Risikobereitschaft des Kreditgebers verknüpft, während j komplementär sein Verlustrisiko beschreibt, wenn die Unsicherheit U aufgeteilt wird. Obige Formeln erlauben es in diesem Fall die Risiken genauer zu quantifizieren.

5 Fazit

Das wesentliche Ergebnis der vorliegenden Arbeit lässt sich wie folgt zusammenfassen:

Im Rahmen der Kreditvergabe bietet der DS-Ansatz einem risikobereiten Kreditgeber quantifizierte Maße für das erweiterte Gewinnpotenzial und die zugehörigen Ausfallwahrscheinlichkeiten an. Beide Größen werden auf die Unsicherheit U zurückgeführt. Ein risikobereiter Investor mit höherer Gewinnabsicht wird akzeptieren, dass Teile des Maßes für die Unsicherheit U in das Wahrscheinlichkeitsmaß $p(G)$ fließen. Dies wird durch die Einführung des Skalierungsfaktors i erreicht. Ein risikobewusster Investor, der Verluste eher scheut, wird hingegen ein Teil der Unsicherheit U dem Wahrscheinlichkeitsmaß $p(B)$ zuschlagen. Dies wird durch den zweiten Skalierungsfaktor j beschrieben. Beide Faktoren sind nicht unabhängig voneinander, da immer gilt $0 \leq i + j \leq 1$.

Schlussendlich ist die Anwendung der DS-Theorie für risikobereitere Kreditgeber neben der klassischen Wahrscheinlichkeitstheorie eine durchaus interessante Methode zur Ermittlung der Bonität eines potenziellen KNs. Zudem erweist sie sich als eine an die Risikobereitschaft des Kreditgebers flexibel anpassbare Theorie.

Literatur

- [1] Ballal, Prasanna: *Dempster Shafer Theory*, Intelligent Control Systems, 2004
- [2] Berndt, Jan Ole: *Modellierung von Unsicherheit in der Agenten-Programmierung*, Universität Bremen, 2006
- [3] Bundesverband deutscher Banken: *Kredit-Scoring - Bestandteil der modernen Kreditvergabe*, Bankenverband - Bundesverband deutscher Banken, Berlin, Oktober 2006, Seite 9
- [4] Curley, Shawn P.: *The application of Dempster-Shafer theory demonstrated with justification provided by legal evidence*, University of Minnesota, 2007
- [5] Dempster, Arthur. P.: *New methods for reasoning towards posterior distributions based on sample data.*, Harvard University Seiten 355-374, 1966
- [6] Dempster, Arthur. P.: *Upper and lower probabilities induced by a multivalued mapping*, Harvard University, Seiten 325-339, 1967
- [7] Dempster, Arthur. P.: *A generalization of Bayesian inference*, Harvard University and London School of Economics, Seiten 205-247, 1968
- [8] Prof. Dr. Göbel Stefan: *Formale Abbildung widersprüchlicher Evidenz im Kontext der Dempster-Shafer-Theorie*, Universität Rostock, 2011
- [9] Khaleghi, Bahador; Khamis, Alaa; Karray, Fakhreddine O., and Razavi, Saiedeh N.: *Multisensor data fusion: A review of the state-of-the-art*, Information Fusion, Seiten 28-44, 2011
- [10] Dr.Oberkampff, William L: *Uncertainty Quantification: Using Evidence Theory*, Stanford University, 2005
- [11] Orponen, Pekka: *Dempster's rule of combination is #P complete*, Artificial Intelligence, 44(1-2): Seiten 245-253, 1990
- [12] Parsons, Simon; *Some qualitative approaches to applying the Dempster-Shafer theory*, London, 1994

- [13] Reineking, Thomas: *Belief Functions: Theory and Algorithms*, Universität Bremen , 2014
- [14] Shafer Glenn: *A mathematical theory of evidence*, Princeton University Press, 1976
- [15] Smets, Philippe: *Belief functions*. In Smets, P., Mamdani, E. H., Dubois, D., and Prade, H., editors, *Non Standard Logics for Automated Reasoning*, Seiten 253-286. Academic Press, London, 1988
- [16] Smets, Philippe: *The concept of distinct evidence*. In *Information Processing and Management of Uncertainty*, Seiten 89-94, Université Libre de Bruxelles Brussels-Belgium, 1992
- [17] Thomas Lyn C.: *Consumer Credit Models: Pricing, Profit, and Portfolios* Oxford University Press, New York, 2009.
- [18] Yager, Ron: *On the Dempster-Shafer Framework and New Combination Rules*, Seiten 93-137, Iona College, New Rochelle, New York, 1987
- [19] Zadeh, Lotfi A.: *Fuzzy sets. Information and Control*, Seiten 338-353, University of California, Berkeley, California, 1965.
- [20] Zadeh, Lotfi A.: *On the validity of Dempster's rule of combination of evidence. Technical report, Electronics Research Laboratory*, University of California, Berkeley, California, 1979
- [21] Zadeh, Lotfi A.: *Review of a mathematical theory of evidence*, AI Magazine, Seite 81, University of California, Berkeley, California, 1984