



Universität Augsburg
Prof. Dr. Hans Ulrich Buhl
Kernkompetenzzentrum
Finanz- & Informationsmanagement
Lehrstuhl für BWL, Wirtschaftsinformatik,
Informations- & Finanzmanagement

UNA
Universität
Augsburg
University

Diskussionspapier WI-192

Der effiziente Einsatz proaktiver und reaktiver Datenqualitätsmaßnahmen - ein modellbasierter Ansatz und seine Anwendung bei einem Finanzdienstleister

von

Bernd Heinrich

in: Die Betriebswirtschaft 67 (2007) 5, S. 539-562

**Der effiziente Einsatz proaktiver und reaktiver Datenqualitätsmaßnahmen –
ein modellbasierter Ansatz und seine Anwendung bei einem Finanzdienstleister**

Bernd Heinrich

Anschrift:

Dr. Bernd Heinrich

Universität Augsburg, Lehrstuhl für Betriebswirtschaftslehre, Wirtschaftsinformatik &
Financial Engineering, Universitätsstrasse 16, 86135 Augsburg, bernd.heinrich@wiwi.uni-augsburg.de

Der effiziente Einsatz proaktiver und reaktiver Datenqualitätsmaßnahmen – ein modellbasierter Ansatz und seine Anwendung bei einem Finanzdienstleister

Stichworte: Datenqualität, Datenqualitätsmaßnahmen, Finanzdienstleister

Zusammenfassung:

In vielen Großunternehmen werden jährlich Millionen für Maßnahmen zur Verbesserung der Datenqualität (DQ) ausgegeben, wobei überwiegend in reaktive Datenqualitätsmaßnahmen (DQM) investiert wird (beispielsweise die Prüfung auf Adressfehler mit Hilfe von Assoziationsregeln). Hierbei werden oftmals die Größe einer existierenden Datenbasis und deren schlechte DQ als Entscheidungskriterien für eine DQ-Investition angeführt. Sie lassen sich jedoch - so zeigen die Ergebnisse des Beitrags - nicht bestätigen. Vielmehr muss wegen der Charakteristik der reaktiven DQM beispielsweise je Datenattribut und Assoziationsregel entschieden werden, ob diese einen positiven Grenzbeitrag leisten. Entsteht ein solcher Beitrag nicht, so sind selbst bei einer großen, zu sichernden Datenmenge reaktive DQM nicht zu ergreifen. Daneben sind die heute noch stark vernachlässigten proaktiven DQM wegen ihres Multiplikatoreffekts den reaktiven DQM unter sonst gleichen Voraussetzungen sogar vorzuziehen. Proaktive DQM setzen unmittelbar bei der Datenerfassung an (zum Beispiel im Kundengespräch) und erlauben somit nach der Qualitätssicherung die Daten direkt wieder für eine bessere Beratung zu nutzen. Jedoch bergen DQM, insbesondere proaktive Maßnahmen, aufgrund ihrer speziellen, zeitverzögerten Wirkungen die Gefahr einer Fehlinvestition. Große DQ-Projekte mit der in der Praxis oftmals genannten Zielsetzung schnell eine hohe Qualität realisieren zu wollen, sind unter diesem Aspekt nicht sinnvoll.

Summary:

Large companies spend huge amounts of money to improve their data quality. Our approach points out that these investments are much too high if certain special effects of proactive data quality measures are not considered. From this point of view, expensive data quality projects with the objective to provide high data quality in one step, are not reasonable. In addition, proactive measures of data quality (like improved processes of data recording and transformation) are under the same conditions preferable, compared to today's commonly used reactive measures (for example associativity rules). Another result concerns the fact that a company has a lot of customer data which are not quality-checked: the literature recommends the implementation of reactive data quality measures first. Here we show that a company has to decide on each data attribute and associativity rule whether (or not) they add a positive marginal contribution. In case there is no such contribution, reactive measures should not be taken even if the customer data that need to be quality-checked are huge.

Der effiziente Einsatz proaktiver und reaktiver Datenqualitätsmaßnahmen – ein modellbasierter Ansatz und seine Anwendung bei einem Finanzdienstleister

1 Einleitung

Der erfolgreiche Einsatz und die Akzeptanz von Informationssystemen in den Bereichen Business Intelligence, Managementunterstützung oder Customer Relationship Management (CRM) hängt sehr stark davon ab, ob eine Integration und Veredelung von Datenbeständen aus operativen Prozessen, beispielsweise mittels Data-Warehouse-Systemen, gelingt. Eine Problemstellung, die hierbei sowohl in der Forschung als auch in der Praxis intensiv diskutiert wird, ist die Gewährleistung und Verbesserung der Datenqualität (DQ). DQ ist deshalb so wichtig, da sowohl (operative) Geschäftsprozesse als auch Managementprozesse Daten als Durchführungs- und Entscheidungsgrundlage benötigen. Können dabei Qualitätskriterien wie Vollständigkeit, Korrektheit, Konsistenz oder Aktualität nicht gewährleistet werden, so wirkt sich dies auf viele Bereiche, wie beispielsweise die Güte von Managemententscheidungen (DQ als Hauptansatzpunkt zur Verbesserung von Business Intelligence Funktionen [1]) oder die Vertriebs- und Marketingkosten bei Kundenkampagnen (zum Beispiel aufgrund von Adressfehlern [2]) negativ aus. Insofern wird sowohl in wissenschaftlichen [3] als auch in praxisorientierten Beiträgen [4] diskutiert, welcher Nutzen durch Datenqualitätsmaßnahmen (DQM) entsteht und welche Kosten durch schlechte DQ verursacht werden. Zahlen besagen, dass der Anteil am Budget von Data-Warehouse-Projekten für geplante und vor allem ungeplante DQM wegen inkorrektur und fehlender Daten bei bis zu 50% liegt [5]. Auch auf der Fachseite führt eine schlechte DQ zu erheblichen Problemen. So gaben beispielsweise in einer Umfrage der SAS Institute GmbH 67% beziehungsweise 76% der Marketingleiter an, dass die Zufriedenheit ihrer Kunden beziehungsweise die Profitabilität von Kampagnen infolge schlechter DQ nachhaltig leide [6]. DQ wird demnach auch als kritischer Faktor bei der Einführung von CRM-Lösungen gesehen [7]. Die Frage, in welchem Umfang in DQM investiert werden soll und inwiefern der angestrebte Nutzen durch die ergriffenen Maßnahmen überhaupt erreicht wird, wird jedoch nicht untersucht.

Daher wird im Beitrag die Fragestellung aufgegriffen, wie stark unter Kosten-/Nutzensgesichtspunkten in DQM investiert werden soll, wobei das Einsatzgebiet des CRM - speziell die Kundenberatung - zugrunde gelegt wird. Der derzeitige Stand der Literatur ist im Kapitel 2 dargestellt. Darauf aufbauend wird in Kapitel 3 ein Modell entwickelt, welches die Kostenwirkungen von DQM und deren Nutzenwirkung infolge einer verbesserten Kundenbeziehung aufzeigt. Dabei wird, im Vergleich zu bisherigen Ansätzen, zwischen reaktiven und proaktiven DQM unterschieden. Reaktive Maßnahmen, wie zum Beispiel Data

Cleansing-Maßnahmen, besitzen analytischen Charakter (Qualitätssicherung „im nachhinein“) und wirken sich auf die DQ bereits *vorhandener* Kundendaten aus. Proaktive Maßnahmen, welche die Verbesserung der Datenerhebungs- und Transformationsprozesse betreffen, zeichnen sich dagegen durch ihren konstruktiven Charakter aus und beeinflussen die Qualität *zukünftig zu erfassender* Daten. Die Anwendung des Modells an einem Fallbeispiel wird in Kapitel 4 ausführlich erläutert. Im abschließenden Kapitel 5 werden die wesentlichen Ergebnisse festgehalten.

2 Analyse und Diskussion bisheriger Beiträge

Die Konsequenzen einer schlechten Qualität von in Geschäfts- und Managementprozessen verwendeten Daten und die resultierenden ökonomischen Folgen werden derzeit intensiv diskutiert (für einen Überblick vergleiche [8]). Der Ausgangspunkt der Überlegungen ist dabei, die Kosten- und Nutzeneffekte möglicher DQM gegenüber zu stellen und auf dieser Basis eine (optimale) Entscheidung über den Umfang der zu ergreifenden Maßnahmen zu treffen. Für den hier fokussierten Bereich des CRM werden beispielsweise in [9] die (modelltheoretischen) Grundlagen gelegt, indem Wirkungszusammenhänge zwischen DQM und der Verbesserung der Kundenbeziehung untersucht werden. Eine verbesserte Beziehung spiegelt sich hier in Form einer Erhöhung der Kundentransaktionen (innerhalb eines Zeitraums) wider, wobei Transaktionen aufgrund ihrer besseren Messbarkeit (beispielsweise im Vergleich zur Kundenzufriedenheit [10]) Verwendung finden. Dadurch ist die Wirkung der DQM ex post auch besser zu erheben. Zudem wird in [9] von einem begrenzten, verfügbaren Transaktionsvolumen des Kunden (zum Beispiel das Vermögen des Kunden zur Geldanlage) ausgegangen, das dieser zwischen verschiedenen Unternehmungen (beispielsweise verschiedene Finanzdienstleister zur Geldanlage) aufteilen kann. Als Zielgröße bei der Planung der DQM wird dementsprechend der Transaktionsanteil $\lambda \in [0;1]$ einer Unternehmung (mit Summe der Transaktionsanteile über alle Unternehmungen gleich eins) und dessen Veränderung $\Delta\lambda$ infolge einer höheren DQ verwendet.

Die hierbei durchgeführten Analysen verdeutlichen, dass es erst ab einem bestimmten Umfang an DQM (so genanntes Signifikanzniveau) und der damit erzielten DQ zu einer Verbesserung der Kundenbeziehung kommt. Dies ist dadurch begründet, dass ein Kunde erst entsprechende Abwanderungs- oder Ausstiegsbarrieren der Konkurrenz (zum Beispiel bedingt durch Kosten wie Auflösungsgebühren oder nicht realisierte Prämien), aber gegebenenfalls auch die eigene Trägheit oder Bequemlichkeit überwinden muss, bevor es zu einer Steigerung des Transaktionsanteils λ kommen kann. Im Folgenden wird dies vereinfacht durch den

Zusammenhang $\lambda = f(DQM)$ in der Form $\lambda = m \cdot DQM^\eta$ veranschaulicht, das heißt der Anteil λ hängt vom Umfang der ergriffenen DQM (mit $DQM \in [0; 1]$) ab. Die Parameter m und η besitzen dabei in Abhängigkeit vom Kunden(segment) unterschiedliche Ausprägungen und geben an, wie ein Kunde(nsegment) beispielsweise die bessere Kenntnis über die eigene Person infolge korrekter und vollständiger Daten bei einer Beratung schätzt. Da erste DQM vom Kunden stärker wahrgenommen werden als eine Ausweitung bereits umfangreicher Maßnahmen, ergibt sich ein konkaver Verlauf der Funktion $\lambda = f(DQM)$ mit $\eta \in]0; 1[$. Wegen $\lambda \leq 1$ gilt zudem $m \cdot DQM^\eta \leq 1$. Zudem muss bei einer derartigen Interpretation die ceteris paribus-Klausel berücksichtigt werden, das heißt die zugrunde gelegte Kundendatenmenge sowie alle anderen Nutzen- und Kostengrößen der Transaktionstätigkeit bleiben unverändert.

In Abb. 1 werden die Wirkung der DQM sowie das Signifikanzniveau an einem exemplarischen Funktionsverlauf verdeutlicht (hier wären durchaus auch mehrere Signifikanzniveaus und damit Unstetigkeitsstellen denkbar). Die DQM entfalten dabei erst ab einem Umfang von $DQM = 0,36$ (Signifikanzniveau) ihre Nutzenwirkung und führen in diesem Fall zu einem Anstieg des bisherigen Transaktionsanteils der Unternehmung von $\lambda_{alt} = 0,3$ auf $\lambda_{sig} = 0,5$ ($\Delta\lambda = 0,2$). Unterhalb des Signifikanzniveaus bleiben die Maßnahmen dagegen aus den oben genannten Gründen wirkungslos, das heißt die Kundenbeziehung würde sich für diesen Fall nicht verbessern.

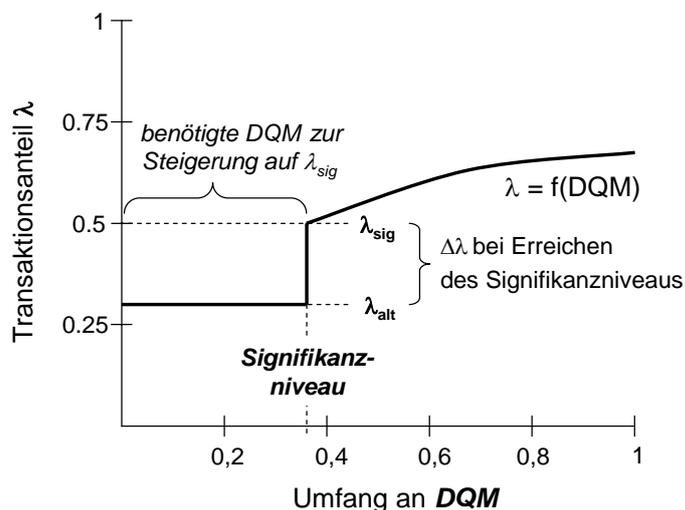


Abb. 1 Auswirkungen von DQM auf die Transaktionstätigkeit des Kunden

Neben dem Signifikanzniveau wird in [9] auch eine Wirkung der DQM aufgezeigt, die Multiplikatoreffekt genannt wird. Dieser Effekt ist darin begründet, dass durch den erhöhten Transaktionsanteil und die damit einhergehende Interaktionstätigkeit mit dem Kunden wieder neue und aktuelle Daten erhoben werden können. Diese führen ihrerseits durch Überprüfung, Vervollständigung und Aktualisierung der bereits vorhandenen Datenbasis zu einer Steigerung

der DQ, die in nachfolgenden Interaktionen wiederum dazu genutzt werden kann, einen höheren Nutzen für den Kunden zu kreieren (beispielsweise kundenindividuellere Leistungsangebote). Resultat ist somit ein erneut positiver Effekt auf den Transaktionsanteil λ (das heißt DQ und Transaktionsanteil beeinflussen sich wechselseitig positiv über die Kundendatenmenge), wobei die DQM hierzu jedoch nicht in ihrem Umfang erhöht werden müssen. Dieser Vorgang wird als Multiplikatoreffekt bezeichnet, der in dieser Weise bei anderen Instrumenten zur Verbesserung der Kundenbeziehung, wie beispielsweise bei monetären Anreizen (Bonus oder Rabatt) nicht auftritt [11].

Aufbauend auf diesen Wirkungszusammenhängen wird in [12] eine mehrperiodige ökonomische Analyse durchgeführt, die jedoch nur reaktive DQM berücksichtigt. Somit fehlt die wichtige Differenzierung verschiedener Maßnahmentypen, insbesondere zwischen proaktiven und reaktiven Maßnahmen. Reaktive DQM besitzen analytischen Charakter und wirken sich dementsprechend auf die DQ - vor allem die Konsistenz - bereits intern *vorhandener* oder gegebenenfalls extern zugekaufter Kundendaten aus. An erster Stelle ist hier an Assoziationsregeln zu denken. Diese beschreiben immer oder häufig auftretende Zusammenhänge zwischen Datenattributwerten durch Regeln, wie beispielsweise den einfachen logischen Ausdruck $a \text{ AND } b \Rightarrow c$ (anwendbar zum Beispiel zur Ermittlung von Adressfehlern ($ORT \text{ AND } STRASSE \text{ AND } HAUSNR \Rightarrow PLZ$)). Generiert beziehungsweise angewendet werden können Assoziationsregeln daneben auch auf extern erworbene Daten. So kann eine erweiterte Kontrolle der Konsistenz der Daten durch ein verbessertes Regelwerk erfolgen, falls der Datenlieferant die Qualität, beispielsweise von Adressen, zusichert. Daneben können externe Daten - auch ohne Verwendung von Assoziationsregeln - als Basis für Korrektheitsprüfungen dienen. Demgegenüber sind proaktive DQM durch ihren konstruktiven Qualitätssicherungscharakter gekennzeichnet, da sich diese - durch Verbesserung der Erhebungs- und Transformationsprozesse - auf die Korrektheit und Vollständigkeit *zukünftig zu erfassender* Daten auswirken. Da gerade proaktive Maßnahmen einen Ansatzpunkt darstellen, um in der eigenen Datenbank nicht zunächst falsche (Kunden-)Daten zu speichern und diese anschließend wieder mühsam bereinigen zu müssen, sind diese natürlich von besonderer Bedeutung. Beispiele für proaktive DQM sind IT-gestützte Hilfen für Mitarbeiter bei der Dateneingabe (um beispielsweise auf inkonsistente Eingaben beziehungsweise Begriffe hinzuweisen und diese teilautomatisiert zu korrigieren), Metadatenmanagement, Normalisierung der verwendeten Datenschemata in den Quellsystemen oder der Abgleich sowie die Integration von Datenschemata vor der Durchführung einer Datenmigration. Wie derartige proaktive DQM jedoch modelltheoretisch überhaupt berücksichtigt werden können wird in [12] nicht diskutiert. Demzufolge kann mit den dort vorgestellten Investitionsmustern und Handlungsempfehlungen auch die Fragestellung, inwieweit nunmehr in proaktive respektive in reaktive Maßnahmen zu investieren ist - das heißt, ob eine wechselseitige Beeinflussung (zum Beispiel Verstärkung) der

Wirkung der Maßnahmentypen beziehungsweise eine Substituierbarkeit existiert - nicht beantwortet werden.

Gerade dies erscheint jedoch vor dem Hintergrund der Bedeutung proaktiver DQM und der Tatsache, dass sich über den Einsatz proaktiver und reaktiver DQM bislang nur wenige Aussagen in der Literatur finden lassen, besonders wichtig. So postuliert beispielsweise English in diesem Zusammenhang, dass *meist* einmalige Bereinigungsaktionen (reaktive DQM) bei der initialen Einführung eines DQ-Managements geeignet sind [13]. Werden demgegenüber die Daten, die zu qualitätssichern sind, in *häufig* durchgeführten Prozessen generiert beziehungsweise dafür genutzt, so sind die Gründe mangelnder DQ zu identifizieren und daraus Maßnahmen zur Sicherstellung und kontinuierlichen Verbesserung (proaktive DQM) der DQ umzusetzen. Dabei lassen Worte wie *meist* und *häufig* bereits auf die schwierige Fundierung dieser Aussagen schließen, da weder die Rahmenbedingungen der Entscheidung noch die Wirtschaftlichkeit der DQM betrachtet werden. Letzteres spiegelt sich ebenso in Beschreibungen praktischer Anwendungen wider, wenn reaktive DQM bei großen Konsistenzproblemen bestehender Datenbasen und proaktive DQM bei komplexen Erfassungsprozessen vorgeschlagen werden [14].

Nachstehend ist das Fallbeispiel eines Finanzdienstleisters beschrieben, das die Problemstellung nochmals verdeutlichen soll und zugleich als Basis für die Operationalisierung und Anwendung des Ansatzes in Kapitel 4 dient.

Ein FDL will zur individualisierten Beratung im Segment der gehobenen Privatkunden („mass affluent“ Segment) mit einem frei verfügbaren Anlagevermögen von 100-200 T€ ein Kundenmodell einführen. Das Kundenmodell entspricht einem Datenrepository [15], in dem Kunden- und Transaktionsdaten strukturiert gespeichert werden (die Abb. 2 zeigt das vereinfachte Datenschema). Hierzu gehören unter anderem Daten über die familiären und beruflichen Verhältnisse, die Kundeneinstellungen, wie zum Beispiel die Risikobereitschaft und Einschätzungen über den Kunden, wie beispielsweise die Risikotragfähigkeit (wie viel Risiko sollte ein Kunde aufgrund seines Einkommens und Vermögens eingehen?). Vorteile resultieren aus der Möglichkeit einer automatisierten Auswertbarkeit der Daten im Kundenmodell und einer dadurch möglichen weit reichenden IT-Unterstützung des Beraters oder gar einer Selbstberatung des Kunden mit Hilfe des IT-Systems. Beides ist nicht nur vor dem Hintergrund einer Individualisierung zu sehen, sondern ebenfalls bezüglich Haftungsfragen des FDL infolge einer stark schwankenden Beratungsqualität (abhängig von den Fähigkeiten und dem Kenntnisstand eines Kundenberaters).

Neben dem fachlichen Design des Kundenmodells wird ebenfalls die Frage der Qualitätssicherung der Kundendaten aufgeworfen. Konkret soll beantwortet werden, in welchem Umfang DQM generell und proaktive und reaktive DQM im Speziellen bei Einführung des Kundenmodells zu ergreifen sind.

Als reaktive DQM fokussiert der FDL primär die vorgenannten Assoziationsregeln. Auf Basis der Analyse des Datenschemas (inhaltliche Abhängigkeiten einzelner Attribute) oder des Datenbestands (statistisch ermittelte Abhängigkeiten) werden derartige Regeln ermittelt und können nach ihrer semantischen Prüfung als Integritätsbedingungen für den Bestand dienen (Details vergleiche Anhang 3). Demgegenüber wird als proaktive DQM die Verbesserung der Erfassungsprozesse der Kundendaten betrachtet. Hier kann bei den Dateneingaben die dem Kundenmodell zugrunde liegende semantische Modellierung [16] genutzt werden. Die Realweltbegriffe, welche für die Erfassung notwendig sind, werden inhaltlich beschrieben und bereitgestellt (beispielsweise in sehr einfacher Weise innerhalb einer Drop-down-Box), um unabhängig vom Datenerfasser eine *einheitliche* Begriffsverwendung zu ermöglichen.

Die DQM sind dabei natürlich in ihrem Umfang (zum Beispiel Menge von Assoziationsregeln) variierbar, womit sich die konkrete Frage stellt (auch zur Konfiguration eines DQ-Tools, vergleiche Anhang 3), in welchem Umfang in proaktive und reaktive DQM unter ökonomischen Gesichtspunkten zu investieren ist. Unter ökonomischen Aspekten deswegen, da die DQM nicht aus einem Selbstzweck heraus durchgeführt werden sollen.

3 Optimierungsmodell zur Ermittlung des effizienten Umfangs an Datenqualitätsmaßnahmen

Im Weiteren wird ein Modell zur Beantwortung obiger Investitionsfragestellung entwickelt. Dabei muss im Ansatz zwischen proaktiven und reaktiven Maßnahmen unterschieden werden, um die jeweilige unterschiedliche Kosten-/Nutzenwirkung modellbasiert berücksichtigen zu können. Die Ausgestaltung und Operationalisierung der beiden Maßnahmentypen durch spezifische DQM erfolgt danach in der praktischen Anwendung beim FDL.

Folgende Annahmen und Definitionen werden für das Modell zugrunde gelegt [17]:

- (A1) Es wird ein einperiodiges Entscheidungsmodell mit den Unterperioden t (mit $t = 1, 2, \dots$) betrachtet. Dabei trifft die Unternehmung nur zu Beginn der Periode eine Entscheidung über den Umfang der reaktiven (DQM_R) beziehungsweise proaktiven Maßnahmen (DQM_P) und setzt diese ohne Zeitverzögerung um. In den nachfolgenden Unterperioden t treten die Nutzen- und Kostenwirkungen der DQM ein, wobei keine weiteren Entscheidungen bezüglich der Maßnahmen getroffen werden. Die Maßnahmenintensitäten DQM_P und DQM_R sind auf den Bereich $[0; 1]$ normiert. Dabei bedeutet $DQM_P, DQM_R = 0$, dass jeweils keine proaktiven oder reaktiven

Maßnahmen durchgeführt werden, wohingegen bei $DQM_p, DQM_R = 1$ die Intensitäten maximal sind (vergleiche hierzu die Operationalisierung in Kapitel 4 und im Anhang 3) [18].

- (A2) Ein Kunde tätigt bei der Unternehmung einen Anteil λ seines gesamten Transaktionsvolumens (mit $\lambda \in [0; 1]$). Das Volumen berücksichtigt dabei auch das transaktionsorientierte, noch nicht realisierte Kundenpotenzial [19]. Da die Höhe des Anteils λ durch DQM beeinflussbar ist, setzt sich λ aus dem DQ-induzierten Anteil λ_{DQM} und einem auch ohne DQM zu Stande kommenden, gegebenen Anteil λ_{-DQM} zusammen ($\lambda_{DQM,t}, \lambda_{-DQM} \in [0; 1]$ mit $\lambda_{DQM,t} + \lambda_{-DQM} = \lambda \leq 1$).
- (A3) Die Menge der Kundendaten der Unternehmung resultiert vollständig aus der Transaktions- und Interaktionstätigkeit mit dem Kunden. Externe Datenquellen (vor allem Datenzukauf) werden nicht berücksichtigt. Die Menge der Kundendaten setzt sich aus der zu Beginn der Periode vorhandenen initialen Datenmenge D_0 und den bis zur Unterperiode t insgesamt neu hinzukommenden Daten D_t zusammen. Da die Mengen D_0 und D_t den Anteil an den Kundendaten gemäß Spezifikation (beispielsweise für das Datenschema in Abb. 2) wiedergeben und damit einem spezifikationsorientierten DQ-Begriff [20] folgen, zielen beide auf das Qualitätskriterium der Vollständigkeit der Datenbasis ab. Deshalb wird $D_0 + D_t$ (für alle Unterperioden t) auf den Bereich $[0; 1]$ normiert. Dabei bedeutet $D_0 + D_t = 0$, dass keine der in der Spezifikation geforderten Kundendaten vorhanden sind. Dagegen sind bei $D_0 + D_t = 1$ alle benötigten Daten laut Spezifikation vorhanden. Daneben können die jeweiligen Kunden unterschiedliche Transaktionsvolumina besitzen. Daher existiert ein exogen vorgegebenes $\tau \in [0; 1]$, welches festlegt, wie viele Daten ein Kunde bei der Durchführung des gesamten Transaktionsvolumens maximal (in Bezug auf die Datenspezifikation) weitergibt, das heißt es gilt: $D_0 + D_t \leq \tau \forall t$.
- (A4) Die Unternehmung verfolgt das Ziel der Maximierung des qualitätsinduzierten Nettonutzens $E \in \mathfrak{R}$, das heißt der Differenz zwischen dem Bruttonutzen $U \in \mathfrak{R}^+$, resultierend aus dem DQ-induzierten Anteil λ_{DQM} abzüglich der für die DQM entstehenden Kosten $K \in \mathfrak{R}^+$. Für die Nettonutzenfunktion E sollen hier zunächst weder das Signifikanzniveau beim Nutzen noch sprungfixe Kosten (im Inneren des Definitionsbereichs von DQM_p und DQM_R) betrachtet werden.
- (A5) Es steht zunächst ein unbegrenztes Budget zur Realisierung der DQM zur Verfügung [21].

(A6) Für DQM_p und DQM_R , die Kundendaten D_0 und D_t sowie für λ_{DQM} und λ_{-DQM} gilt jeweils die Eigenschaft der beliebigen Teilbarkeit.

Unter der obigen Annahme (A4) ergibt sich folgende Zielfunktion:

$$E(DQM_p, DQM_R) = U(DQM_p, DQM_R) - K(DQM_p, DQM_R) \rightarrow \max! \quad (1)$$

Der Bruttonutzen U entsteht dabei durch den insgesamt DQ-induzierten Transaktionsanteil λ_{DQM} :

$$U(DQM_p, DQM_R) = u \cdot \lambda_{DQM} \quad (2)$$

Der Faktor $u \in \mathbb{R}^+$ entspricht hier demjenigen Nutzen(faktor), den die Unternehmung nach Abzug aller sonstigen Kosten – außer den DQ-Kosten – bei Erhöhung des Anteils λ_{DQM} von null auf eins realisiert. Er kann aus der internen Rechnungslegung [22] ermittelt werden und soll wegen der Betrachtung des einzelnen Kunden und den damit einhergehenden begrenzten Auswirkungen auf die Struktur der Leistungserstellung über den gesamten Definitionsbereich von λ_{DQM} konstant sein. Um den qualitätsinduzierten Anteil λ_{DQM} zu generieren, sind DQM durchzuführen. Dabei ist jedoch zwischen proaktiven und reaktiven DQM zu unterscheiden.

Reaktive Maßnahmen DQM_R stellen eine „ex post Qualitätssicherung“ dar. Da die reaktiven Maßnahmen zu Beginn der Periode realisiert werden, sind sie auf die zu diesem Zeitpunkt schon vorhandene initiale Datenmenge D_0 anzuwenden. Demgegenüber sichern die proaktiven Maßnahmen DQM_p die in den Unterperioden t neu hinzukommenden Daten D_t . Letztere resultieren einerseits aus den Transaktionen, die der Kunde auch ohne DQM tätigt (vergleiche λ_{-DQM}). Zum anderen kann bei Umsetzung der DQM durch eine bessere Kundeninteraktion zusätzlich ein Anteil λ_{DQM} erzielt werden. Durch diese zusätzlichen Transaktionen wird in einem nächsten Schritt (in der nächsten Unterperiode) die neu hinzukommende Datenmenge gesteigert [23]. Die Daten können mit den proaktiven DQM wiederum qualitätsgesichert werden, wodurch sie erneut zu einer verbesserten Kundeninteraktion beitragen. Dieser rekursive Effekt wird durch den in Kapitel 2 diskutierten Multiplikatoreffekt der DQ beschrieben. Damit ergibt sich der bis zur Unterperiode t insgesamt generierte qualitätsinduzierte Anteil $\lambda_{DQM,t}$ wie folgt:

$$\lambda_{DQM,t} = m \cdot (DQM_p^{\eta_P} \cdot D_t + DQM_R^{\eta_R} \cdot D_0) \quad (3)$$

mit : $D_t = f(\lambda_{-DQM} + \lambda_{DQM,t-1})$

Der Term (3) verdeutlicht, dass sich der Anteil $\lambda_{DQM,t}$ als Ergebnis des (einmaligen) Einsatzes von DQM_R auf die Datenmenge D_0 und dem Einsatz von DQM_p auf die insgesamt neu

hinzukommende Datenmenge D_t bestimmen lässt. Dabei beschreiben die Exponenten η_R und η_P den Wirkungsgrad der jeweiligen DQM. Sie geben wie der Parameter m (mit $m \in [0;1[$) an, wie ein Kunde(nsegment) die korrekten Daten bei der Interaktion – vor allem in der Beratung – schätzt (Operationalisierung vergleiche Kapitel 4) [24]. Da DQ-Verbesserungen vom Kunden bei einer geringen DQ in größerem Umfang wahrgenommen werden als weitere Steigerungen einer bereits existierenden hohen DQ, gilt $\eta_P, \eta_R \in [0;1[$ (abnehmende Grenznutzenwirkung).

Die bis zur Unterperiode t insgesamt neu hinzukommende Datenmenge D_t resultiert – wegen der Nichtberücksichtigung externer Datenquellen (vergleiche Annahme (A3)) – aus der Transaktionstätigkeit, das heißt aus dem Anteil λ_{-DQM} und dem Anteil $\lambda_{DQM,t-1}$ in der Unterperiode $t-1$. Die Unterperiode $t-1$ wird wegen des Multiplikatoreffekts herangezogen, da die DQM_P die Korrektheit der in den Unterperioden erfassten Daten (infolge der Transaktionsdurchführung) verbessert. Hier soll im Weiteren für $D_t = f(\lambda_{-DQM} + \lambda_{DQM,t-1})$ vereinfacht $D_t = \tau \cdot (\lambda_{-DQM} + \lambda_{DQM,t-1})$ gelten. Dabei legt $\tau \in]0; 1]$ fest (vergleiche Annahme (A3)), wie viele Daten ein (spezifischer) Kunde in Bezug auf die Spezifikation infolge der gesamten Transaktionsdurchführung maximal weitergibt.

Demnach lässt sich (3) unter Berücksichtigung des Multiplikatoreffekts schreiben als:

$$\lambda_{DQM,t} = m \cdot DQM_P^{\eta_P} \cdot \tau \cdot (\lambda_{-DQM} + \lambda_{DQM,t-1}) + m \cdot DQM_R^{\eta_R} \cdot D_0 \quad (3')$$

Wegen $\lambda_{DQM,t} = f(\lambda_{DQM,t-1}, \dots)$ entspricht (3') einer Differenzgleichung. Diese lässt sich analog für alle Unterperioden t darstellen. Wie in Anhang 1 gezeigt wird, konvergiert (3') und damit der Multiplikatoreffekt. Der Konvergenzpunkt ergibt sich zu:

$$\lambda_{DQM} = \frac{m \cdot (DQM_P^{\eta_P} \cdot \tau \cdot \lambda_{-DQM} + DQM_R^{\eta_R} \cdot D_0)}{1 - m \cdot DQM_P^{\eta_P} \cdot \tau} \quad (4)$$

Der Grenzwert λ_{DQM} in (4) für den DQ-induzierten Anteil ist zentral für die weiteren Überlegungen. Er beschreibt den gesuchten Zusammenhang zwischen den ergriffenen Maßnahmen DQM_P und DQM_R und dem DQ-induzierten Anteil, wobei im Vergleich zu Kapitel 2 der Multiplikatoreffekt nunmehr explizit berücksichtigt ist. Mit (4) lässt sich (2) wie folgt darstellen:

$$U(DQM_P, DQM_R) = u \cdot \lambda_{DQM} = u \cdot \left(\frac{m \cdot (DQM_P^{\eta_P} \cdot \tau \cdot \lambda_{-DQM} + DQM_R^{\eta_R} \cdot D_0)}{1 - m \cdot DQM_P^{\eta_P} \cdot \tau} \right) \quad (2')$$

Neben der Nutzenfunktion (2') ist auch die Kostenfunktion K , die den bewerteten Ressourceneinsatz für $DQM_p, DQM_R \in]0; 1]$ ausdrückt, zu spezifizieren:

$$K(DQM_p, DQM_R) = K_p + k_p \cdot DQM_p^{\mu_p} \cdot D_i + K_R + k_R \cdot DQM_R^{\mu_R} \cdot D_0 \quad (5)$$

Zum einen steigen hier die Qualitätskosten mit dem Umfang der ergriffenen Maßnahmen DQM_p und DQM_R an, wobei k_p und k_R die jeweiligen Kostenfaktoren darstellen. Darüber hinaus werden mit den Parametern μ_p beziehungsweise μ_R die steigenden Grenzkosten (mit $\mu_p, \mu_R > 1$) für einen zunehmenden Umfang der DQM berücksichtigt. So ist der Ressourceneinsatz zur Steigerung einer geringen DQ kleiner als wenn bei einer schon hohen DQ noch weitere Verbesserungen durch Maßnahmen erzielt werden sollen. Beispiele hierfür sind Data Cleansing-Maßnahmen, die bei ihrer initialen Anwendung (zum Beispiel Bereinigung offensichtlicher Adressfehler) mit wenig Ressourceneinsatz zu realisieren sind. Weniger offensichtliche Mängel sind dagegen erst mit komplexen Assoziationsregeln zu beseitigen, in deren Entwicklung weit mehr investiert werden muss. Die Grenzkosten verhalten sich damit analog zu den vorgestellten *appraisal* und *prevention costs*.

Zum zweiten werden die DQM auf die jeweilige Kundendatenmenge angewendet. Bei DQM_p bedeutet dies, dass zum Beispiel qualitätssichernde Erfassungs- und Transformationsprozesse bei einer größeren, neu hinzukommenden Datenmenge D_i stärker beansprucht werden, weswegen auch der entsprechende Ressourceneinsatz steigt. Demgegenüber ist DQM_R nur auf die initiale Datenmenge D_0 anzuwenden.

Zum dritten werden neben maßnahmenvariablen Kosten auch fixe Bestandteile $K_p, K_R \in \mathfrak{R}^+$ für $DQM_p, DQM_R > 0$ berücksichtigt. Diese entfallen deswegen für $DQM_p = 0$ beziehungsweise $DQM_R = 0$, da die Konzeption und Durchführung von DQM Projektcharakter besitzen, das heißt falls keine DQM ergriffen werden, sind auch etwaige Fixkosten des DQ-Projektes nicht zu berücksichtigen. Bei den Fixkosten handelt es sich primär um Kosten für Bestandsaufnahme, Anforderungsspezifikation und Festlegung der DQ-Rahmenbedingungen (beispielsweise organisatorische Einbettung, Rollen und Verantwortlichkeiten). Die Höhe ist dabei erfahrungsgemäß nicht zu unterschätzen, weswegen die Randlösungen, bei denen die fixen Kostenbestandteile K_p und/oder K_R entfallen, ebenso zu untersuchen sind.

Wie der Ausdruck (5) zeigt, ist bei den Kosten der proaktiven Maßnahmen die neu hinzukommende Datenmenge zu berücksichtigen, die ihrerseits (wegen

$$D_i = \tau \cdot (\lambda_{-DQM} + \lambda_{DQM,t-1})) \text{ vom DQ-induzierten Anteil abhängt. Insofern wirkt sich der}$$

Multiplikatoreffekt der proaktiven Maßnahmen auch bei den Kosten aus. Somit ergibt sich:

$$\begin{aligned}
K(DQM_p, DQM_R) &= \\
&= K_p + k_p \cdot DQM_p^{\mu_p} \cdot \tau \cdot (\lambda_{-DQM} + \lambda_{DQM}) + K_R + k_R \cdot DQM_R^{\mu_R} \cdot D_0 \\
&= K_p + k_p \cdot DQM_p^{\mu_p} \cdot \tau \cdot (\lambda_{-DQM} + \left(\frac{m \cdot (DQM_p^{\eta_p} \cdot \tau \cdot \lambda_{-DQM} + DQM_R^{\eta_R} \cdot D_0)}{1 - m \cdot DQM_p^{\eta_p} \cdot \tau} \right)) + K_R + k_R \cdot DQM_R^{\mu_R} \cdot D_0
\end{aligned} \tag{5'}$$

Auf Basis von (2'), (4) und (5') kann die Nettonutzenfunktion E nunmehr ermittelt werden, die hier (und generell unter der Annahme (A4)) im Inneren des Definitionsbereichs von DQM_p und DQM_R zweimal stetig differenzierbar ist. Für die Bestimmung der Optima von E sind die Ableitungen $\delta E / \delta DQM_p$ beziehungsweise $\delta E / \delta DQM_R$ zu bilden, umzuformen und jeweils Null zu setzen. Der Anhang 2 enthält die Ableitungen wie auch die Diskussion der Krümmungseigenschaften der Nettonutzenfunktion. Es lässt sich unter der hinreichenden Bedingung (A') zeigen, dass die Nettonutzenfunktion konkav ist und somit die nachstehenden Ausdrücke (6) und (7) zu einem globalen Maximum im Definitionsbereich $DQM_p, DQM_R \in]0; 1]$ führen. Der Ausdruck (6) zeigt das Optimum für DQM_R :

$$DQM_{R,opt} = \left(\frac{m \cdot \eta_R \cdot (u - DQM_p^{\mu_p} \cdot k_p)}{k_R \cdot \mu_R \cdot (1 - m \cdot DQM_p^{\eta_p})} \right)^{\frac{1}{\mu_R - \eta_R}} \tag{6}$$

Das Optimum für DQM_p kann dagegen nur als implizite Funktion (7) angegeben werden:

$$DQM_{p,opt}^{\eta_p} \cdot m \cdot u \cdot \eta_p - DQM_{p,opt}^{\mu_p} \cdot k_p \cdot (DQM_{p,opt}^{\eta_p} \cdot m \cdot (\eta_p - \mu_p) + \mu_p) = 0 \tag{7}$$

Darüber hinaus sind für $DQM_p = 0$ beziehungsweise $DQM_R = 0$ jeweils die Randlösungen zu prüfen.

Bevor die Interpretation der Modellergebnisse und die Operationalisierung erfolgen, ist in Bezug zur Annahme (A4) noch kurz auf das Signifikanzniveau einzugehen. Liegen die ermittelten, optimalen Werte für DQM_p und DQM_R unterhalb des Signifikanzniveaus (vergleiche Abb. 1), so wird dadurch *keine* Steigerung des Transaktionsanteils ausgelöst. Wegen der konkaven Nettonutzenfunktion im Definitionsbereich von DQM_p und DQM_R ist jedoch zu folgern:

Grundsätzlich gilt, dass ausgehend vom Optimum bei einer weiteren Steigerung der DQM der Nettonutzen monoton abnimmt. Wenn diese Steigerung jedoch notwendig ist, um das Signifikanzniveau zu überwinden, ist es zweckmäßig denjenigen Umfang an DQM_p und DQM_R zu prüfen, mittels dessen gerade das Signifikanzniveau und damit das korrespondierende λ_{sig} erreicht werden kann. Eine darüber hinausgehende Erhöhung von DQM_p beziehungsweise DQM_R ist nicht sinnvoll, da dies den Nettonutzen weiter reduzieren

würde. Die effiziente Kombination von DQM_P und DQM_R zur Erreichung von λ_{sig} ist dabei unter Berücksichtigung der proaktiven und reaktiven Wirkungs- und Kostenfaktoren zu ermitteln (vergleiche hierzu den Anhang 4 für ein Beispiel). Analog zu oben, sind dabei auch die beiden Randlösungen noch zu prüfen. Wechselbarrieren von Kunden, die in der praktischen Anwendung zu einem Signifikanzniveau führen, können somit Berücksichtigung finden.

4 Anwendung des Modells und Interpretation der Ergebnisse

Zur Operationalisierung des Modells wird das obige Fallbeispiel des FDL fortgeführt. Wie in Kapitel 2 geschildert, wird das Kundenmodell zur Anlageberatung im Segment der gehobenen Privatkunden eingesetzt. Um hierbei den Umfang zu ergreifender DQM zu bestimmen, sind in einem ersten Schritt die einzelnen Modellparameter als Durchschnitt des Kundensegments zu erheben beziehungsweise abzuschätzen (aus Anonymisierungsgründen wurden die Zahlen hierbei verändert).

Als Anteil λ_{-DQM} wird dabei der \emptyset -Anteil des vom FDL verwalteten Vermögens am gesamten geschätzten Anlagevolumen des Kundensegments (vor Durchführung der DQM) verwendet. Der Anteil λ_{-DQM} wird von der Fachseite mit durchschnittlich 33% angegeben [25], das heißt nur etwa $\frac{1}{3}$ des Vermögens der Kunden wird derzeit bei der Unternehmung angelegt ($\frac{2}{3}$ bei Konkurrenten). Der Bruttonutzen u für eine Steigerung des Anlageanteils kann aus dem bisherigen Grenzbeitrag je 1000 € zusätzlichem Anlagevermögen des zuständigen Geschäftsbereichs kalkuliert werden (Deckungsbeitragsrechnung), wobei hier die Kosten der zu ergreifenden DQM natürlich nicht berücksichtigt sind. Im Beispiel wird $u=19$ (in Mio. €) veranschlagt, das heißt kann der durchschnittliche Anteil infolge von DQM um 1% gesteigert werden, so ergibt sich ein Grenzbeitrag von 190 T€ zur Deckung der DQM-Kosten.

Für die Erhöhung des Anlageanteils will die Unternehmung bei den reaktiven DQM Assoziationsregeln einsetzen, die mit dem Frequent-Pattern-Growth Algorithmus gebildet werden (insgesamt wurden noch weitere spezifische DQM mit dem Ansatz auf ihren Einsatz hin untersucht und in der Wirkung miteinander verglichen) [26]. Dieses datenorientierte Verfahren berechnet die Häufigkeit von 1-Itemmengen in Datenbanken (vergleiche Anhang 3). Über den Mindestsupport wird dabei angegeben, wie häufig eine 1-Itemmenge in der Datenbasis auftreten muss, damit sie als Assoziationsregel, zum Beispiel $a \text{ AND } (b \text{ XOR } c) \Rightarrow d$ in Frage kommt. Demgegenüber wird mit der Mindestkonfidenz die Untergrenze für die relative Häufigkeit festgelegt, mit der bei Auftreten der Aussage $a \text{ AND } (b \text{ XOR } c)$ zugleich auch die Aussage d (und keine andere Aussage) als Implikation folgt. Mit beiden Parametern kann damit stufenlos

und toolbasiert der Umfang möglicher Assoziationsregeln und damit die Höhe von DQM_R umgesetzt werden.

Demgegenüber will die Unternehmung bei den proaktiven DQM für die Datenerfassung standardisierte Begriffe vorgeben, die dem Berater inhaltlich erläutert werden (Bedeutung und Verwendungskontext wird vorgegeben) und bei der Dateneingabe in den Feldern zur Verfügung stehen. Dadurch soll eine konsistente Begriffsverwendung erreicht werden, auf deren Basis die Kundencharakteristik (beispielsweise welche Risikoeinstellung besitzt ein Kunde?) automatisiert und unabhängig vom Verständnis des einzelnen Beraters ermittelt werden kann.

Die Basis der Qualitätssicherung bildet dabei das Kundenmodell. Seine Spezifikation erfolgt in einem fachlichen Datenschema. Die Abb. 2 zeigt eine verkürzte Darstellung:

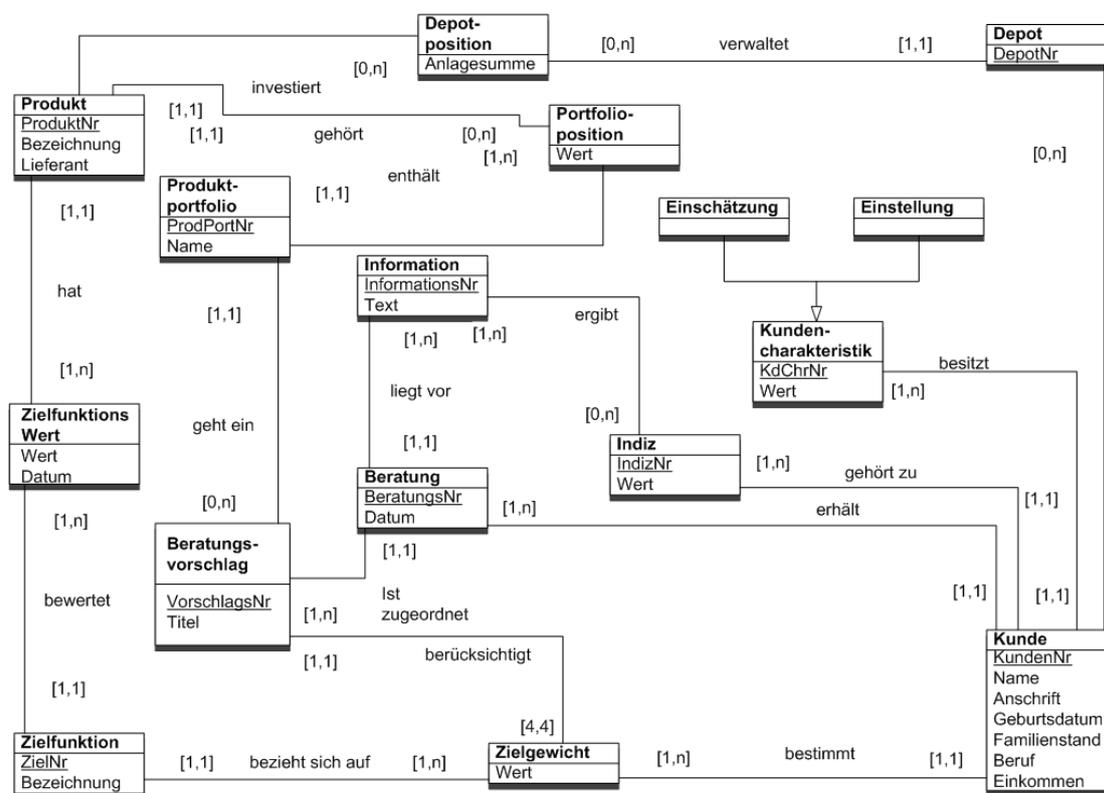


Abb. 2 Vereinfachtes Datenschema des Fallbeispiels

Mittels des Schemas kann eine individualisierte Beratung unterstützt werden, indem aus den erfassten Kundeninformationen (zum Beispiel Geburt des ersten Kindes eines Kunden), Indizien zur Ableitung der Kundencharakteristika ermittelt werden [27]. Letztere dienen beispielsweise zur Einschätzung der Risikoneigung (Verringerung der Risikobereitschaft durch die Geburt des ersten Kindes) oder der Flexibilitätsansprüche des Kunden (zum Beispiel hinsichtlich Liquidität des Anlageprodukts) und werden bei der Erstellung des Beratungsvorschlags (zum Beispiel Austausch von Produkten im Anlageportfolio) verwendet. Das Kundenmodell wird dabei aus Gründen der Gewährleistung einer hohen und mitarbeiterunabhängigen Beratungsgüte

(beispielsweise aufgrund der vorhandenen Produktkomplexität) und der Beraterhaftung eingesetzt.

Spezifikationsorientierte DQ auf Basis des spezifizierten Datenschemas ist hier in zweierlei Hinsicht von Bedeutung. Zum einen in der traditionellen Form durch Überprüfung der Vollständigkeit der Datenattribute (zum Beispiel ist das Datenattribut *Beruf* bei allen Kunden gespeichert). Insofern kann die Datenmenge D durch Gegenüberstellung der vorhandenen zu den benötigten Attributwerten unproblematisch ermittelt werden. Im Anwendungsfall liegen für das Segment bisher 42% der Attributwerte vor, das heißt $D_0 = 0,42$ [28].

Zum anderen hat die DQ direkten Einfluss auf die inhaltliche Ausgestaltung des Beratungsvorschlags und damit auf seine Akzeptanz beim Kunden. Daher sollen potenzielle Widersprüche in den gespeicherten Daten entdeckt und mit dem Kunden abgeklärt werden. So könnte ein Kunde auf Basis seiner bisherigen Transaktionen in relativ risikoreiche Anlageprodukte als wenig risikoavers gelten. Wenn eine hohe Risikobereitschaft jedoch aufgrund der familiären (zum Beispiel Kunde ist Alleinversorger der Familie) und beruflichen (Arbeitsplatz gefährdet) Verhältnisse des Kunden als inadäquat eingestuft wird (Risikotragfähigkeit ist geringer als seine Risikobereitschaft), so kann dies bei der Beratung thematisiert und im Vorschlag berücksichtigt werden. Dementsprechend sind auch die Parameter η_P , η_R und m abzuschätzen. Sie sind ein Wirkungsparameter der DQM für die Generierung des Anteils λ_{DQM} , das heißt es ist zu untersuchen, inwieweit vollständige und korrekte Daten den Beratungsvorschlag aus Kundensicht verbessern und damit den Anlageanteil steigern. Beispielsweise bedeutet dies für die Ermittlung des Wirkungsparameters η_P , im Beratungsgespräch die Daten mit den standardisierten Begriffen zu erfassen, daraufhin die Kundencharakteristik (zum Beispiel Risikotragfähigkeit) besser ermitteln zu können und diese direkt für die Vorschlagserstellung zu verwenden (um hinsichtlich der Beraterhaftung auf die Gefahren von $\text{Risikobereitschaft} > \text{Risikotragfähigkeit}$ hinzuweisen). Um die Wirkung einer derartigen Maßnahme abzuschätzen, ist sie an einer Stichprobe des Kundensegments zu testen. Ein ceteris paribus-Vergleich zwischen den bisherigen Akzeptanzquoten von Beratungsvorschlägen und denjenigen der Stichprobe erlaubt eine Abschätzung des Wirkungsparameters η_P (analog η_R für die Wirkung von Assoziationsregeln). Der Wirkungsparameter der DQM_P wird im Fallbeispiel mit $\eta_P=0,34$ schlechter (wegen $DQM \in [0; 1]$) verringert sich mit steigendem η die Wirkung der DQM) als derjenige der DQM_R mit $\eta_R=0,25$ geschätzt (Auf das Vorgehen und die Güte dieser Schätzung wird später noch eingegangen). Der Lageparameter m ist im Beispiel 0,1.

Daneben sind Kostenschätzungen für die Konzeption und Umsetzung der DQM vorzunehmen. Hier ist einerseits die Höhe der vom Umfang der DQM unabhängigen (fixen) Kosten zu bestimmen, die beispielsweise für Teile der Institutionalisierung des DQ-Projekts anfallen. Im

Beispiel wurden diese mit $K_P=190$ T€ beziehungsweise $K_R=140$ T€ veranschlagt. Andererseits sind die variablen Kostensätze k_P und k_R sowie die Exponenten μ_P und μ_R abzuschätzen.

Beispielsweise können k_R und μ_R ermittelt werden, indem geprüft wird, wie viel Mehrkosten durch die Erhöhung der Anzahl der eingesetzten Assoziationsregeln sowie deren Evaluierung und Anwendung entstehen. Eine hohe Maßnahmenintensität führt dabei zur Berücksichtigung komplexer Regeln, die einen geringeren Support- und Konfidenzwert besitzen und deren Prüfung auf Validität sowie deren Anwendung auf die Datenbasis hohe Kosten verursachen (letzteres wegen notwendiger Plausibilitätstests). Gerade deswegen wurde in Kapitel 3 ein konvexer Verlauf der Kostenkurven argumentiert. Im Fallbeispiel werden $k_P=0,85$ und $k_R=0,81$ sowie $\mu_P=1,4$ und $\mu_R=1,3$ abgeschätzt, das heißt die proaktiven DQM werden im Vergleich zu den reaktiven DQM aufgrund der notwendigen umfangreichen organisatorischen Verankerung im Beratungsprozess mit höheren Kostenparametern versehen.

In welchem Umfang soll der FDL nunmehr reaktive beziehungsweise proaktive DQM ergreifen? Da bereits im Durchschnitt über das Kundensegment 42% der benötigten Attribute vorhanden sind und im Vergleich dazu – infolge des Basistransaktionsanteils – weniger neue Daten hinzukommen, müsste er eigentlich primär auf reaktive DQM setzen (so auch die Empfehlungen der Literatur, vergleiche Kapitel 2). Zudem besitzen die proaktiven DQM im Beispiel eine geringere Wirkung und höhere Kostenparameter. Insofern könnte man folgern, dass in erster Linie DQM_R (Assoziationsregeln) und nur in geringerem Umfang DQM_P zu ergreifen sind. Beispielsweise könnte der FDL DQM_R in Höhe von 0,8 festlegen, das heißt die Mindestkonfidenz und der Mindestsupport würden im Umkehrschluss gering gewählt werden, um auch komplexe, im Datenbestand selten auftretende Assoziationsregeln zu erhalten. Dagegen könnte DQM_P beispielsweise mit 0,1 festgelegt werden, das heißt nur für die wichtigsten Eingabefelder werden Begriffe standardisiert. Falls der Finanzdienstleister, wie beschrieben verfährt, hätte dies folgende Konsequenzen (erste Zeile der Tabelle):

Proaktive Maßnahmenintensität DQM_P	Reaktive Maßnahmenintensität DQM_R	Zus. Transaktionsanteil λ_{DQM}	Zus. Bruttounutzen $u \cdot \lambda_{DQM}$ (in Mio. €)	Proaktive Qualitätskosten $K_P+k_P \cdot DQM_P^{\mu_P} \cdot D_t$ (in Mio. €)	Reaktive Qualitätskosten $K_R+k_R \cdot DQM_R^{\mu_R} \cdot D_o$ (in Mio. €)	Nettonutzen E (in Mio. €)
0,1	0,8	0,057	1,09	0,203	0,395	0,492
0,6	0,5	0,069	1,31	0,355	0,278	0,677

Anhand der Anwendung des Modells ergeben sich dagegen die in Zeile 2 der Tabelle stehenden optimalen Maßnahmenumfänge mit $DQM_{P,opt} = 0,6$ und $DQM_{R,opt} = 0,5$ sowie der Nettonutzen $E = 0,677$ Mio. € (siehe Anhang 4 für die Eigenschaften der Nettonutzenfunktion, die Randlösungen im Beispiel sowie für Ausführungen zum Signifikanzniveau). Die Abb. 3 veranschaulicht für das Beispiel die Entwicklung des Nettonutzens für alternative Umfänge an DQM_P und DQM_R :

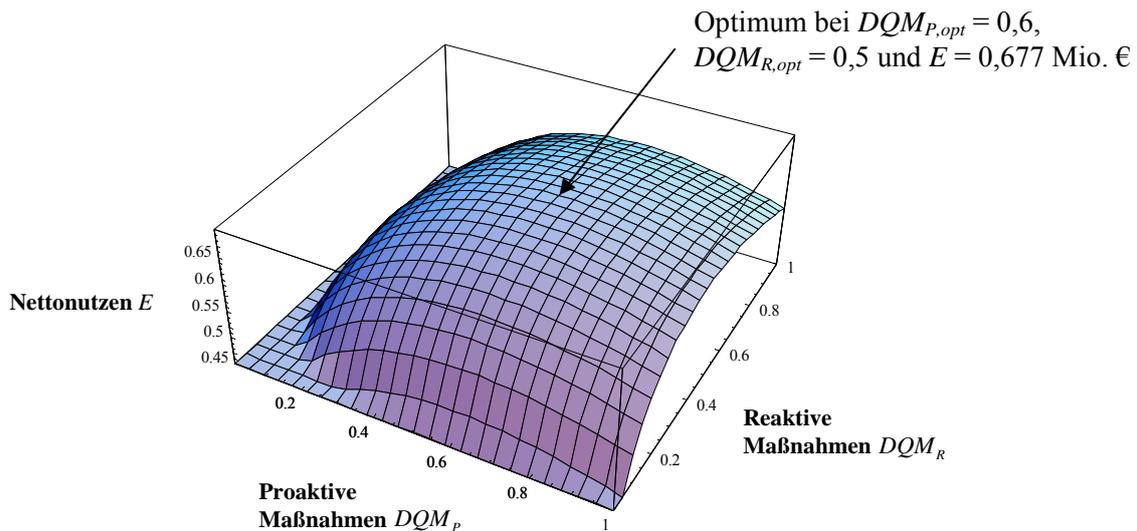


Abb. 3 Darstellung des Nettonutzens im Beispiel

Bei der Analyse der Ergebnisse ist festzustellen, dass $DQM_{P,opt} = 0,6$ höher und $DQM_{R,opt} = 0,5$ geringer als zunächst erwartet ausfallen. Letzteres ist deshalb überraschend, da im Beispiel eine relativ umfangreiche initiale Datenmenge existiert, das heißt die Ergreifung reaktiver Maßnahmen erscheint vorteilhaft gegenüber proaktiven Maßnahmen. Dies trifft jedoch nicht zu, was in der Charakteristik der DQM_R begründet liegt, da letztlich pro Datenattribut und Assoziationsregel – und nicht auf Basis der Größe der gesamten Datenbasis – entschieden werden muss, ob die Regel (noch) einen positiven Grenzbeitrag leistet. Dies würde man per se in den eigenen Überlegungen zunächst nicht berücksichtigen, da auch in der Literatur empfohlen wird, bei einer großen, noch ungesicherten Datenmenge umfangreiche Data Cleansing-Maßnahmen einzusetzen (vergleiche Kapitel 2). Hier ist es jedoch viel sinnvoller und dies lässt sich auch in realen Sachverhalten nachvollziehen, mit gegebenenfalls wenigen DQM – nämlich diejenigen, die einen positiven Grenzbeitrag [29] leisten – eine umfangreiche Datenbasis gezielt zu sichern. Das oftmals angeführte Entscheidungskriterium der Größe einer Datenbasis für den Umfang an DQM lässt sich somit nicht bestätigen.

Analog hierzu sind auch die Modellergebnisse zu verstehen. Die initiale Kundendatenmenge D_0 sowie der initiale Anteil λ_{-DQM} (für $D_0 > 0$ und $\lambda_{-DQM} > 0$) wirken sich auf die Bestimmung des optimalen Umfangs an $DQM_{P,opt}$ und $DQM_{R,opt}$ nicht aus. Sie können beide in den Ableitungen $\delta E / \delta DQM_P$ und $\delta E / \delta DQM_R$ (siehe (2.1) und (2.2) im Anhang 2) ausgeklammert werden und erscheinen deshalb in den Optima (6) und (7) nicht mehr. Dennoch ist anhand der Randlösungen zu überprüfen, ob bei einer geringen Datenmenge D_0 beziehungsweise einem geringen Anteil λ_{-DQM} die verursachten fixen DQM-Kosten überhaupt gedeckt sind, da diese

Größen natürlich die Höhe des Bruttonutzens und der Kosten (vergleiche Terme (2') und (5) beeinflussen.

Neben diesem Resultat ist vor allem auch die Wirkung des Multiplikatoreffekts zu diskutieren, der sich am Beispiel des FDL leicht verdeutlichen lässt. Hier führt der Multiplikatoreffekt dazu, dass trotz der höheren Kostenparameter und des kleineren Wirkungsparameters der DQM_p , diese im Optimum viel stärker als erwartet zu ergreifen sind. Die qualitätsinduzierte Erzielung des Anlageanteils $\lambda_{DQM,t}$ durch die optimalen Maßnahmenumfänge, jedoch noch ohne Multiplikatoreffekt (deshalb $t=1$), lässt sich hier ermitteln mit:

$$\lambda_{DQM,t} = m \cdot DQM_P^{n_P} \cdot (\lambda_{-DQM} + \lambda_{DQM,t-1}) + m \cdot DQM_R^{n_R} \cdot D_0$$

$$\lambda_{DQM,1} = 0,1 \cdot 0,6^{0,34} \cdot (0,33 + 0) + 0,1 \cdot 0,5^{0,25} \cdot 0,42 = 0,063$$

Dabei ist unerheblich, ob die Transaktionen des Anteils λ_{-DQM} gleich zu Beginn oder auf die einzelnen Unterperioden verteilt realisiert werden. Durch die DQM würde demzufolge im Segment der Anlageanteil in der ersten Unterperiode um $\lambda_{DQM,1} = 0,063$ (also mehr als 6% des gesamten Anlagevolumens) erhöhen. Bei der Durchführung dieser Transaktionen werden wiederum Kundendaten erfasst, die wegen der ergriffenen proaktiven Maßnahmen ebenfalls qualitätsgesichert werden. Hier ist beispielsweise daran zu denken, dass durch eine konsistente und semantisch interpretierbare Datenbasis im Kundenmodell individuellere Beratungsvorschläge generiert werden können. In die zweite und jede weitere Unterperiode ist jeweils die Realisierung der Vorperiode einzubeziehen, um $\lambda_{DQM,t}$ zu ermitteln. Es ergeben sich folgende Berechnungen:

$$\lambda_{DQM,2} = 0,1 \cdot 0,6^{0,34} \cdot (0,33 + 0,063) + 0,1 \cdot 0,5^{0,25} \cdot 0,42 = 0,068$$

$$\lambda_{DQM,3} = 0,1 \cdot 0,6^{0,34} \cdot (0,33 + 0,068) + 0,1 \cdot 0,5^{0,25} \cdot 0,42 = 0,069$$

$$\lambda_{DQM,4} = 0,1 \cdot 0,6^{0,34} \cdot (0,33 + 0,069) + 0,1 \cdot 0,5^{0,25} \cdot 0,42 = 0,069$$

Wie der Aufstellung zu entnehmen ist, konvergiert der qualitätsinduzierte Transaktionsanteil auf den Wert 0,069. Der *Multiplikatoreffekt* steigert somit, ohne weitere Erhöhung oder Änderung des Maßnahmenumfangs den Anteil zusätzlich um circa 10%, nämlich von 0,063 auf 0,069.

Verwendet man den Ausdruck (4), so ergibt sich diese Anteilssteigerung direkt mit:

$$\lambda_{DQM} = \frac{0,1 \cdot (0,6^{0,34} \cdot 0,33 + 0,5^{0,25} \cdot 0,42)}{1 - 0,1 \cdot 0,6^{0,34}} = 0,069$$

Die zusätzliche Steigerung durch den Multiplikatoreffekt sieht im ersten Moment gering aus. Wie sich jedoch ein unberücksichtigter Multiplikatoreffekt bei der Planung der DQM auswirkt, lässt sich wie folgt verdeutlichen: Tritt in diesem Zusammenhang die Situation auf, dass seitens des Vertriebs beispielsweise eine qualitätsinduzierte Steigerung des Transaktionsvolumens um

7% vorgegeben wird, so müssten im Fallbeispiel die Umfänge der beiden DQM jeweils auf circa 0,75 (Ergebnis der Optimierung ohne Multiplikatoreffekt) erhöht werden. Insofern würden die DQM viel höher (das heißt DQM_R um 50% und DQM_P um 25% im Vergleich zum Optimum) eingeplant werden, mit dem Resultat einer Verringerung des Nettonutzens um circa 15%. Generell lässt sich daraus zum einen folgern, dass bei einer Nichtberücksichtigung des Multiplikatoreffekts in der betrieblichen Praxis, die DQM als zu hoch und damit ökonomisch nicht sinnvoll angesetzt werden. Zum zweiten betont der Multiplikatoreffekt, der unmittelbar dafür genutzt werden kann mit den neu erfassten, qualitätsgesicherten Daten im Kundengespräch weitere Transaktionen zu initiieren, die Bedeutung der proaktiven Maßnahmen gegenüber den heute überwiegend genutzten Assoziationsregeln (DQM_R). Unter sonst gleichen Voraussetzungen (beispielsweise gleiche Kostenstruktur) sind DQM_P deswegen auch den DQM_R vorzuziehen. Einschränkungen hiervon ergeben sich für den Fall, dass die DQM_P aufgrund des gegebenenfalls tieferen Eingriffs in die Organisation (zum Beispiel veränderter Erfassungs- und Beratungsprozess) erheblich höhere Kosten als die DQM_R verursachen.

Daneben fällt bei den Modellergebnissen auf, dass der Ausdruck für $DQM_{P,opt}$ von Größen der reaktiven Qualitätssicherung (DQM_R , η_R , k_R und μ_R) unabhängig ist (vergleiche Ausdruck (7)). Dies liegt daran, dass die DQM_R zwar die Ausgangsdatenbasis auf Korrektheit und Konsistenz prüfen, die DQM_P jedoch nur die neu hinzukommenden Daten sichern. Ein Zusammenhang im Sinne, je umfangreicher DQM_R ergriffen werden, umso mehr lässt sich auf DQM_P verzichten (Substitutionsargument), lässt sich damit nicht bestätigen. Praktisch ist dies auch für den Einsatz von DQM in einer Periode nachvollziehbar und übertragbar, da die DQM_P die in Vergangenheit erfassten Daten nicht qualitätssichern (für die praktische Umsetzung der optimalen DQM vergleiche Anhang 3).

Neben den diskutierten Ergebnissen, stellt sich beim Ansatz die praktische Schätzung der Parameter η_P , η_R , μ_P und μ_R als kritisch dar. Neben der oben erläuterten Stichprobendurchführung wurde im Fallbeispiel damit so umgegangen, dass nicht Einzelwerte sondern Bereiche für die Parameter abgeschätzt wurden, um so die Fehlertoleranz zu erhöhen. Auf dieser Basis wurden Simulationen mit jeweils 300 Durchläufen ausgeführt, bei denen per Zufallsstichprobe Werte aus den Bereichen gezogen wurden. Damit wurde ermittelt, ob die optimale Umfänge an $DQM_{P,opt}$ und $DQM_{R,opt}$ – bezogen auf die Parameterbereiche – stark streuen (Sensitivitätsanalyse bei Veränderung mehrerer Parameter). Die Abb. 4 zeigt für die Intervalle $\eta_P \in [0,26; 0,42]$, $\eta_R \in [0,17; 0,33]$, $\mu_P \in [1,32; 1,48]$, $\mu_R \in [1,22; 1,38]$, $k_P \in [0,77; 0,93]$ und $k_R \in [0,72; 0,88]$ den Bereich der optimalen DQM-Umfänge im Beispiel. Wie die Graphik veranschaulicht, ist das obige Ergebnis mit $DQM_{P,opt} = 0,6$ und $DQM_{R,opt} = 0,5$ relativ stabil.

Das heißt selbst wenn die praktische Schätzung der Parameter mit Fehlern behaftet ist, weichen die Ergebnisse nur gering ab.

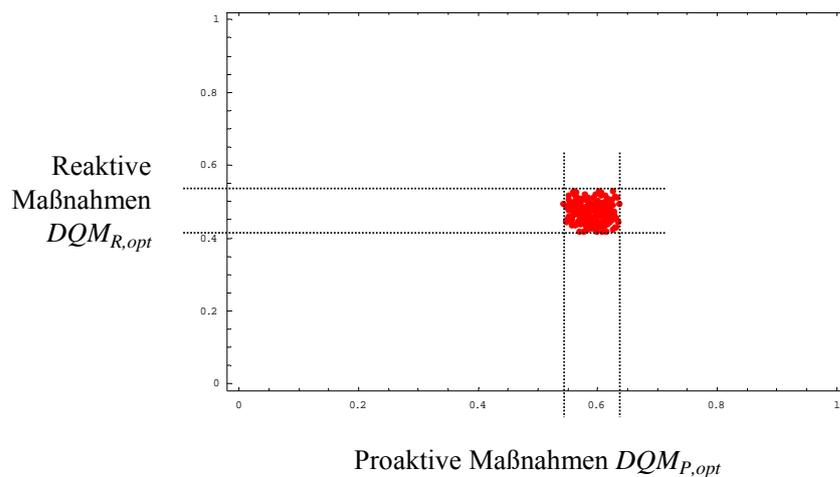


Abb. 4 Robustheit des optimalen Umfangs an DQM im Beispiel

Einen zweiten Kritikpunkt stellt die ex ante Messung der DQ-Wirkung dar. Während die Kosten, die Datenmengen und Transaktionsanteile aber auch der Multiplikatoreffekt (wie viel neue Daten werden durch eine weitere Transaktion erfasst und wie verändern diese die Beratungsvorschläge?) gut abzuschätzen sind [30], liegt das Hauptproblem in der Praxis in der schwierigen Isolierbarkeit und Zurechenbarkeit der Wirkung der DQM (qualitätsinduzierter Anteil $\lambda_{DQM,t}$ der Transaktionen). In einer Unternehmung findet eine Vielzahl von Eingriffen gleichzeitig statt, die den Transaktionsanteil eines Kundensegments verändern. Auf ihn wirken zudem auch nicht zu eliminierende unternehmensexterne Einflüsse (vor allem von Konkurrenten) ein. Dennoch gibt es verschiedene Wege - je nach zu bereinigendem Fehlertyp und der Nutzung der Daten - einer ex ante Ermittlung des qualitätsinduzierten Anteils $\lambda_{DQM,t}$. Handelt es sich zum Beispiel um die Fehlerbereinigung bei Kundenadressen, die zur Unterbreitung von Produktangeboten oder zur Kampagnendurchführung genutzt werden, so lässt sich damit die Quote fehlerhaft angeschriebener Kunden reduzieren. Diese Quote, die in der Praxis oftmals nicht gerade gering ist (bei Kampagnen sind Quoten von ca. 35% keine Ausnahme), kann durch den Einsatz von Assoziationsregeln zur Qualitätssicherung der Adressen (zum Beispiel um 10%-15%) gesenkt werden. Das heißt, liegt die durchschnittliche Erfolgsquote einer Kampagne beispielsweise bei ca. 10% (jeder zehnte, der erreichten Kunden nimmt das Angebot wahr), so könnten die anteiligen Verkaufserfolge der durch die Adresskorrektur zusätzlich erreichten Kunden dem DQM zugerechnet werden. Derartige Überlegungen oder Analysen historischer Daten können als eine Basis zur Abschätzung qualitätsinduzierter Transaktionen dienen. Bei einer Reihe anderer Fehlertypen, wie beispielsweise bei falsch gespeicherten Vertragsdaten, aufgrund derer den Kunden schlechte Produktangebote unterbreitet werden, kann ähnlich verfahren werden. Auch hier ist auf Basis von Erfahrungswerten bisheriger Kampagnen die Wirkung einer Bereinigung von Vertragsdaten

abzuschätzen. Schwieriger gestaltet sich dabei die Ermittlung des qualitätsinduzierten Anteils, falls Fehlertypen verbessert werden, über deren Wirkung keine oder nur wenig Erfahrungswerte vorliegen. Hier könnte zum Beispiel - wie oben beschrieben - die DQ-Wirkung anhand einer Stichprobe abgeschätzt werden, indem die Maßnahmen zunächst nur auf einen Teil des ansonsten (unternehmensintern und -extern) identisch bearbeiteten Kundensegments angewendet werden. Ein Vergleich zwischen den beiden Teilsegmenten hinsichtlich der durchschnittlich hinzugewonnenen Daten und den Anteilserhöhungen kann dann eine Basis für die Ermittlung der DQ-Wirkung sein, wobei auch hier die Problematik der Validität und Reliabilität der Schätzung zu beachten ist.

5 Zusammenfassung und Ausblick

Im Beitrag wurde ein modellbasierter Ansatz entwickelt, mit dessen Hilfe sich der optimale Umfang an zu ergreifenden reaktiven und proaktiven DQM ermitteln lässt. Für die Anwendung und Operationalisierung wurde der Fall eines FDL geschildert. Wesentliche Ergebnisse sind:

- Der in der Literatur angeführte Zusammenhang, nachdem bei einem größeren vorhandenen Datenbestand auch hohe reaktive DQM durchzuführen und zugleich proaktive DQM zweitrangig sind, ist nicht zu bestätigen. Der Grund hierfür ist, dass durch die Charakteristik der reaktiven DQM pro Datenattribut und Assoziationsregel entschieden werden muss, ob diese einen positiven Grenzbeitrag leisten. Das oftmals angeführte Entscheidungskriterium der Größe einer Datenbasis für den Umfang an reaktiven (aber auch proaktiven) DQM ist somit nicht zu bestätigen. Demgegenüber wird der Umfang der DQM viel stärker durch deren Auswirkungen beim Endnutzer (Kunden) beeinflusst, das heißt durch dessen Wertschätzung beispielsweise eines besseren Beratungsvorschlags. Komplexe Assoziationsregeln, ermittelt durch ein DQ-Tool, können zwar zur Qualitätssicherung des Datenbestands beitragen, wenn sich dies jedoch nur wenig auf die (oftmals vernachlässigte) Kundenakzeptanz der Produktvorschläge auswirkt, ist ihr Einsatz unter ökonomischen Aspekten nicht sinnvoll.
- Proaktive DQM sind wegen des Multiplikatoreffekts den reaktiven DQM unter sonst gleichen Voraussetzungen vorzuziehen. Proaktive DQM setzen unmittelbar bei der Datenerfassung an (beispielsweise im Kundengespräch) und erlauben somit nach der Qualitätssicherung die Daten direkt für eine bessere Beratung zu nutzen. Im Fall des FDL bedeutet dies, aktuelle und korrekte Daten für die Ermittlung des Anlagevorschlags auf Basis von Kundeneinstellungen und –einschätzungen direkt wieder zu verwenden.

- Vernachlässigt man den Multiplikatoreffekt, so wird bei Zielvorgaben des Fachbereichs (beispielsweise bestimmte Transaktionsvolumensteigerung) zu viel in DQM investiert, da die zeitverzögerten Wirkungen einer einmal verbesserten DQ unberücksichtigt bleiben. Investitionen in DQM sollten demnach in der Praxis sukzessiv getätigt und jeweils adjustiert werden, da einige positive Wirkungen erst sukzessive eintreten. Kostenintensive Großprojekte mit der oftmals genannten Zielsetzung eine hohe DQ in einem Schritt realisieren zu wollen, sind unter diesem Aspekt ebenfalls nicht sinnvoll.
- Unternehmungen, die für ihre Problemstellung detaillierte Empfehlungen zum Einsatz einzelner DQM benötigen, können den dargestellten Ansatz (mit überschaubarem Aufwand) operationalisieren und anwenden. Damit können spezifische DQM, die im Modell aufgrund der notwendigen Abstraktion nur als Maßnahmentyp berücksichtigt sind, im Anwendungskontext analysiert werden.

Neben den oben genannten kritischen Punkten ist auch das einperiodige Entscheidungsmodell (Annahme (A1)) zu nennen, da in Folgeperioden ebenfalls noch Investitionsentscheidungen getroffen werden können beziehungsweise die in einer Periode aufgebaute DQ natürlich teilweise auch noch Bestand hat. Deshalb wird derzeit an der Erweiterung des Ansatzes um mehrere Planungsperioden – und unter Berücksichtigung beider Maßnahmentypen – sowie an besseren Metriken zur ex ante Schätzung der DQ-Wirkung gearbeitet.

Anmerkungen

- [1] o. V. (2006a)
- [2] Siech (2006)
- [3] Zum Beispiel Ballou et al. (1998); Ballou et al. (1999); Cappiello et al. (2003); Fisher et al. (2003); Parssian et al. (2004)
- [4] Zum Beispiel Gfaller/Reiter (2005); o. V. (2006b)
- [5] Matzer (2004); Alt (2003)
- [6] SAS Institute (2003)
- [7] Friedmann (2004); Heijkers (2003)
- [8] Heinrich/Klier (2006a)
- [9] Heinrich/Helfert (2003a)
- [10] Größen wie beispielsweise Kundenzufriedenheit sind natürlich gleichwohl auch wichtig, wenn es um die Verbesserung einer Kundenbeziehung geht. Jedoch sollen sich derartige Größen – falls sie Ziel einer Unternehmung sind – letztlich auch in einer intensiveren und längeren Beziehung niederschlagen und damit zu einer Erhöhung der Transaktionen (innerhalb des zugrunde gelegten Zeitraums) führen.
- [11] Zur praktischen Ausgestaltung des Effekts vergleiche Heinrich/Helfert (2003b)
- [12] Heinrich/Klier (2006a)
- [13] English (1999)

- [14] *Redman (1996)*
- [15] *Buhl et al. (2003); Buhl et al. (2004)*
- [16] *Buhl et al. (2003)*
- [17] Für den Fall, dass bei der Optimierung ein Kundensegment betrachtet wird, sind die einzelnen Parameter für das Segment zu bestimmen.
- [18] Beispielsweise heißt dies, dass bei $DQM_R = 1$ alle von einem DQ-Tool gefundenen Assoziationsregeln bei einer Mindestkonfidenz > 0 für die Konsistenzprüfung des Datenbestands auch zum Einsatz kommen. Die Mindestkonfidenz gibt die Wahrscheinlichkeit an, mit der ein Zusammenhang in der Datenbasis auftreten muss, um in die Regelmenge aufgenommen zu werden.
- [19] Um das gesamte Transaktionsvolumen für verschiedene Kunden(segmente) zu ermitteln, kann auf Studien und Umfragen zurückgegriffen werden (vergleiche *Heinrich/Klier (2006a)*).
- [20] In der Literatur wird bezüglich einer spezifikationsorientierten DQ allgemein auf die Unterscheidung zwischen Designqualität und Konformitätsqualität verwiesen (beispielsweise *Teboul (1991)*). Designqualität bezieht sich auf die Übereinstimmung zwischen anwenderbezogenen Anforderungen und deren Konkretisierung in einer Spezifikation. Demgegenüber umfasst Konformitätsqualität die Einhaltung dieser Spezifikation durch originäre DQM (im Beitrag fokussiert). In *Heinrich/Helfert (2003a)* wird diese Unterscheidung bereits aufgegriffen und behandelt. Eine Prüfung und Optimierung der Spezifikation hat mit Verfahren der Informationsbedarfsanalyse zu erfolgen. Diese Verfahren weitläufig diskutiert (vergleiche zum Beispiel *Strauch (2002)*). Die Diskussion hier im Beitrag aufzugreifen beziehungsweise die Designqualität ausführlich im Modell zu thematisieren, stellt aus Sicht der Autoren nur einen geringen Mehrwert dar und würde die gezielte Steuerung der unterschiedlichen Aufgaben des DQ-Managements erschweren.
- [21] Der Beitrag zielt auf die Planung beziehungsweise Festlegung des effizienten Umfangs an DQM ab, das heißt es wird die optimale Investitionshöhe in die proaktiven beziehungsweise reaktiven DQM ermittelt. Diese gehen als Input in den Business Case des zugehörigen DQ-Projektantrags ein. Eine Unternehmung entscheidet auf Basis der Business Cases aller Projektanträge dann über das zu realisierende IT-Projektportfolio, das heißt welche Projekte zu realisieren sind. Als Resultat dieser Entscheidung können einzelnen, genehmigten Projekten gegebenenfalls auch nachgelagert Budgetrestriktionen auferlegt werden. Dies würde dann im vorgestellten Modell zu einer Optimierung unter Nebenbedingungen führen, welche im Ergebnis diejenigen DQM vorzieht, die bei gegebenem Mitteleinsatz den höchsten Nutzen generiert. Da dies jedoch einem nachgelagerten Schritt entspricht, wird im Grundmodell auf die Einführung von Budgetrestriktionen verzichtet.
- [22] Zur Ermittlung des Faktors u in der Praxis, muss analysiert werden, wie hoch der Grenzbeitrag beispielsweise eines weiteren Euros bei der Geldanlage des Kunden, das heißt bei der Erhöhung des Anlagevolumens ist. Meist sind solche Werte aus der internen Rechnungslegung bekannt oder lassen sich dort ermitteln beziehungsweise abschätzen.
- [23] Neben den Transaktionsdaten können in der neu hinzukommenden Datenmenge D_i ohne weiteres ebenfalls Interaktionsdaten berücksichtigt werden (vergleiche *Heinrich/Klier (2006a)*).
- [24] Alternativ könnte auch die durch den Parameter m dargestellte Wirkung der proaktiven beziehungsweise reaktiven DQM beim Kundensegment maßnahmenspezifisch modelliert werden.

Demnach wären die Parameter m_R und m_P zu unterscheiden und mit

$$\lambda_{DQM,t} = m_P \cdot DQM_P^{\eta_P} \cdot D_t + m_R \cdot DQM_R^{\eta_R} \cdot D_0$$
 im Modell zu berücksichtigen. Aus

Komplexitätsgründen wurde darauf jedoch zunächst verzichtet.

- [25] Hier wurde das Anlagevolumen eines idealtypischen Kunden des Segments analysiert und abgeschätzt. Dabei wurden bei den Berechnungen kundenspezifische Einflussfaktoren wie zum Beispiel Alter, Beruf und Einkommen berücksichtigt. Als Input für diese Berechnungen dienten zum einen öffentlich zugängliche Studien (wie beispielsweise „Einkommens- und Verbrauchsstichprobe“ des Statistischen Bundesamts) als auch speziell in Auftrag gegebene Auswertungen (zum Beispiel zur durchschnittlichen Einkommens- und Vermögensentwicklung je nach Berufsgruppe, etc.). In einem zweiten Schritt wurde die derzeitige Zusammensetzung des Kundensegment analysiert, das heißt wie viel Prozent beispielsweise der 45 und 50 Jahre alten Beamten im höheren Dienst etc. im Segment existieren, um diesen jeweils ein geschätztes Kundenpotenzial auf Basis der Berechnungen zuzuordnen. So konnte das noch nicht vom FDL ausgeschöpfte Kundenpotenzial durch einen Vergleich mit dem tatsächlich beim FDL angelegten Vermögen aufgezeigt werden (share-of-wallet).
- [26] *Han et al.* (2000); *Soler/Yankelevich* (2001)
- [27] detailliert in *Buhl et al.* (2004)
- [28] Der Faktor τ wird hier im Beitrag vereinfachend mit eins festgelegt, da sich dadurch die Ergebnisinterpretation nicht ändert.
- [29] In der praktischen Anwendung würde man hier aufgrund der später noch angesprochenen Messproblematik diejenigen DQM realisieren, die einen höheren Grenzbeitrag besitzen.
- [30] Zur ex ante Schätzung und ex post Messung der DQ-Wirkung vergleiche *Heinrich/Klier* (2006b).

Verzeichnis der zitierten Literatur

- Alt, G.* (2003), Sehr geehrter Frau Müller – Falsche Daten sind nicht nur peinlich, sondern verursachen auch hohe Kosten, in: FAZ Nr. 244 vom 21.10.2003, S. B2.
- Ballou, D. P./Tayi, G. K.* (1999), Enhancing data quality in data warehouse environments, in: Communications of the ACM 42 (1999) 1, S. 73-78.
- Ballou, D. P./Wang, R./Pazer, H./Tayi, G. K.* (1998), Modeling Information Manufacturing Systems to Determine Information Product Quality, in: Management Science 44 (1998) 4, S. 462-484.
- Buhl, H. U./Fridgen, M./Volkert, S.* (2003), Systemunterstützt individualisierte Kundenansprache in der Mehrkanalwelt der Finanzdienstleistungsbranche – Repräsentation der Einstellungen von Kunden in einem Kundenmodell, in: Uhr, W./ Esswein, W./Schoop, E. (Hrsg.): Wirtschaftsinformatik 2003, Band II, Heidelberg, S. 201-229.
- Buhl, H. U./Heinrich, B./Steck, W./Winkler, V.* (2004), Konzept zur individualisierten Finanzdienstleistungsberatung für Privatkunden und dessen prototypische Umsetzung, in: Wirtschaftsinformatik, 46 (2004) 6, S.427-438.
- Cappiello, C./Francalanci, Ch./Pernici, B./Plebani, P./Scannapieco, M.* (2003), Data Quality Assurance in Cooperative Information Systems: A multi-dimensional Quality Certificate, in: Catarci, T. (Hrsg.): International Workshop on Data Quality in Cooperative Information Systems, Siena, S. 64-70.
- English, L.* (1999), Improving Data Warehouse and Business Information Quality. New York.
- Friedmann, K.* (2004), BI: Richtige Informationen zur richtigen Zeit, in: Computerwoche Nr. 20 vom 14.05.2004.
- Fisher, C. W./Chengalur-Smith, I./Ballou, D. P.* (2003), The Impact of Experience and Time on the Use of Data Quality Information in Decision Making, in: Information Systems Research, 14 (2003) 2, S. 170–188.
- Gfaller, H./Reiter, M.* (2005), Die Datenqualität misst sich an ihrem wirtschaftlichen Nutzen. In: Computerzeitung Online, abgerufen unter www.computerzeitung.de/O/50/Y/85460/default.aspx am 13.11.2005.
- Han, J./Pei, J./Yin, Y.* (2000), Mining Frequent Patterns without Candidate Generation, in: Proceedings of the 2000 ACM-SIGMOD International Conference on Management of Data. Dallas, S. 1-12.
- Heijkers, M.* (2003), Ohne Datenhygiene kein CRM, in: Computerwoche Nr. 41 vom 10.10.2003.

- Heinrich, B./Helfert, H. (2003a)*, Nützt Datenqualität wirklich im CRM? – Wirkungszusammenhänge und Implikationen, in: Uhr, W./Esswein, W./Schoop, E. (Hrsg.): Wirtschaftsinformatik 2003, Band II, Heidelberg, S. 231-253.
- Heinrich, B./Helfert, H. (2003b)*, Analyzing Data Quality Investments in CRM - a model based approach, in: Proceedings of the 8th International Conference on Information Quality: Cambridge, S. 80-95.
- Heinrich, B./Klier, M. (2006a)*, Ein Optimierungsansatz für ein fortlaufendes Datenqualitätsmanagement und seine praktische Anwendung bei Kundenkampagnen, in: Zeitschrift für Betriebswirtschaft (ZfB), 76, 6, 2006, S. 559-587.
- Heinrich, B./Klier, M. (2006b)*, Metriken und Messverfahren zur Bewertung der Datenqualität als Basis für ein ökonomisch orientiertes Datenqualitätsmanagement, Diskussionspapier des Lehrstuhls WI-IF der Universität Augsburg. Augsburg.
- Matzer, M. (2004)*, Datenqualität frisst die Hälfte des Data-Warehouse-Etats, in: Computer Zeitung Nr. 3 vom 12.01.2004, S. 12.
- o. V. (2006a)*, Auf die Datenqualität kommt es an, in: Computerwoche vom 17.07.2006, S. 8.
- o. V. (2006b)*, Unternehmen fehlt Datenqualitäts-Strategie, in: Computer Zeitung Nr. 39 vom 25.09.2006, S. 12.
- Parssian, A./Sarkar, S./Jacob, V. S. (2004)*, Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product, in: Management Science 50 (2004) 7, S. 967-982.
- Redman, Th. C. (1996)*, Data Quality for the Information Age. Norwood.
- SAS Institute (2003)*, Europäische Unternehmen leiden unter Profitabilitätseinbußen und niedriger Kundenzufriedenheit durch schlechte Datenqualität, Studie der SAS Institute GmbH, Heidelberg.
- Siech, D. K. (2006)*, Fehlertolerante Suche erhöht Datenqualität, in: Computer Zeitung, 8 (2006), S. 23.
- Shank, J. M./Govindarajan, V. (1994)*, Measuring the cost of quality: A strategic cost management perspective, in: Journal of Cost Management, 8 (1994) 2, S. 5-17.
- Soler, S. V./Yankelevich, D. (2001)*, Quality Mining: A Data Mining Method for Data Quality Evaluation, in: Proceedings of the 6th International Conference on Information Quality: Cambridge, S. 162–172.
- Strauch, B. (2002)*, Entwicklung einer Methode für die Informationsbedarfsanalyse im Data Warehousing, Bamberg.
- Teboul, J. (1991)*, Managing Quality Dynamics, New York.

Anhang 1: Konvergenz des Multiplikatoreffekts

Behauptung:

Die Folge $\lambda_{DQM,t} = m \cdot DQM_P^{\eta_P} \cdot \tau \cdot (\lambda_{-DQM} + \lambda_{DQM,t-1}) + m \cdot DQM_R^{\eta_R} \cdot D_0$ für $t=1, 2, \dots$

konvergiert (mit $\lambda_{DQM,1} = m \cdot DQM_P^{\eta_P} \cdot \tau \cdot (\lambda_{-DQM}) + m \cdot DQM_R^{\eta_R} \cdot D_0$), wobei dies gleichbedeutend mit der Aussage ist, dass der Multiplikatoreffekt konvergiert.

Beweis der Behauptung:

$$\lambda_{DQM,t} = m \cdot DQM_P^{\eta_P} \cdot \tau \cdot (\lambda_{-DQM} + \lambda_{DQM,t-1}) + m \cdot DQM_R^{\eta_R} \cdot D_0$$

$$\lambda_{DQM,t} = m \cdot \underbrace{(DQM_P^{\eta_P} \cdot \tau \cdot \lambda_{-DQM} + DQM_R^{\eta_R} \cdot D_0)}_{:= a} + m \cdot \underbrace{DQM_P^{\eta_P} \cdot \tau}_{:= b} \cdot \lambda_{DQM,t-1}$$

$$\lambda_{DQM,t} = a + b \cdot \lambda_{DQM,t-1}$$

Dabei gilt:

1. a und b sind konstant in t;
2. $0 \leq a$;
3. $0 \leq b < 1$;

Es wird die Konvergenz der Folge $\lambda_{DQM,t}$ gezeigt, indem nachgewiesen wird, dass sie

1) monoton wachsend und 2) nach oben beschränkt ist.

zu 1) Nachweis durch Induktion:

Induktionsanfang: $\lambda_{DQM,1} = a + b \cdot 0 \leq a + b \cdot \lambda_{DQM,1} = \lambda_{DQM,2}$

Induktionsannahme: $\lambda_{DQM,t-1} \leq \lambda_{DQM,t}$

Induktionsschluss: $\lambda_{DQM,t+1} = a + b \cdot \lambda_{DQM,t} \geq a + b \cdot \lambda_{DQM,t-1} = \lambda_{DQM,t}$

\Rightarrow 1) gilt

↑
Induktions-
annahme

zu 2)

$$\begin{aligned}\lambda_{DQM,t} &= a + b \cdot \lambda_{DQM,t-1} = a + b \cdot (a + b \cdot \lambda_{DQM,t-2}) = a + b \cdot a + b^2 \cdot (a + b \cdot \lambda_{DQM,t-3}) = \dots = \\ &= \sum_{i=0}^{t-1} b^i \cdot a = a \cdot \sum_{i=0}^{t-1} b^i < a \cdot \underbrace{\sum_{i=0}^{\infty} b^i}_{\substack{\text{Geometrische Reihe} \\ \text{konvergiert wg. } 0 \leq b < 1}} < \infty\end{aligned}$$

⇒ 2) gilt

Wegen des Nachweises von 1) und 2) folgt, dass $\lambda_{DQM,t}$ und somit der Multiplikatoreffekt konvergiert.

Anhang 2: Optimum und Konkavität der Nettonutzenfunktion (mit $\tau = 1$)

Zur Ermittlung des Optimums der Nettonutzenfunktion sind die Ableitungen $\delta E / \delta DQM_P$ und $\delta E / \delta DQM_R$ zu bilden, umzuformen und jeweils Null zu setzen (Ausdrücke (1.1) und (1.2)):

$$\frac{\delta E}{\delta DQM_R} = \left(\frac{DQM_R^{\eta_R} \cdot D_0 \cdot m \cdot (DQM_P^{\eta_P} \cdot k_P - u) \cdot \eta_R - DQM_R^{\eta_R} \cdot D_0 \cdot k_R \cdot (-1 + DQM_P^{\eta_P} \cdot m) \cdot \mu_R}{DQM_R \cdot (-1 + DQM_P^{\eta_P} \cdot m)} \right) = 0 \quad (1.1)$$

$$\frac{\delta E}{\delta DQM_P} = \left(\frac{\overbrace{(\lambda_{-DQM} + DQM_R^{\eta_R} \cdot D_0 \cdot m)}^{>0} \cdot (DQM_P^{\eta_P} \cdot m \cdot u \cdot \eta_P - DQM_P^{\eta_P} \cdot k_P \cdot (DQM_P^{\eta_P} \cdot m \cdot (\eta_P - \mu_P) + \mu_P))}{DQM_P \cdot (-1 + DQM_P^{\eta_P} \cdot m)^2} \right) = 0 \quad (1.2)$$

Der Term (1.1) lässt sich zum Optimum für DQM_R und somit zum Ausdruck (6) umformen (mit Nenner $DQM_R \cdot (-1 + m \cdot DQM_P^{\eta_P}) < 0$). Hingegen lässt sich das Optimum für DQM_P in Ausdruck (7) nur als implizite Funktion angeben. Es entspricht dem zweiten Term im Zähler von Ausdruck (1.2), da für den ersten Term $\lambda_{-DQM} + DQM_R^{\eta_R} \cdot D_0 \cdot m$ (für $D_0 > 0$, da der Einsatz von DQM_R für den Fall, dass keine Datenmenge existiert, zwecklos ist) sowie für den Nenner $DQM_P \cdot (-1 + DQM_P^{\eta_P} \cdot m)^2$ jeweils > 0 gilt.

Zur Bestimmung der Art der Extremstelle wird das Krümmungsverhalten der Nettonutzenfunktion im Definitionsbereich untersucht. Damit $DQM_{P,opt}$ und $DQM_{R,opt}$ ein globales und kein lokales Maximum darstellen, muss die Nettonutzenfunktion im Definitionsbereich streng *konkav* sein. Wegen der Komplexität der 2. Ableitungen der Nettonutzenfunktion werden die Krümmungseigenschaften ihrer Bestandteile analysiert. Die Nettonutzenfunktion besteht aus der Bruttonutzenfunktion (2') abzüglich der Kostenfunktion (5). In Funktion (2') ist der Grenzwert (4) für den Anteil λ_{DQM} zentral, da (2') eine lineare Transformation von (4) und somit das Krümmungsverhalten beider Terme identisch ist. Zur Ermittlung der Krümmungseigenschaften von (4) wird das Hauptminorenverfahren herangezogen.

Die Funktion (4): $f(DQM_P, DQM_R)$ ist für $(DQM_P, DQM_R) \in [0;1] \times [0;1]$ streng *konkav*, wenn das Hauptminorenkriterium erfüllt ist, das heißt

- für den ersten Eintrag der Hesse-Matrix H gilt

$$H_{11} f(DQM_P, DQM_R) = \frac{\partial^2 f}{\partial DQM_P^2} (DQM_P, DQM_R) < 0 \quad \text{mit}$$

$$(DQM_P, DQM_R) \in [0;1] \times [0;1] \quad \text{und}$$

- für die Determinante der Hesse-Matrix H gilt $\det(H) > 0$ mit

$$(DQM_P, DQM_R) \in [0;1] \times [0;1].$$

Die Hesse-Matrix $H f(DQM_P, DQM_R)$ ist von folgender Form:

$$\left(\begin{array}{cc} \frac{DQM_R^{-2+\eta_R} \cdot D_0 \cdot m \cdot (-1 + \eta_R) \cdot \eta_R}{1 - m \cdot DQM_P^{\eta_P}} & \frac{DQM_P^{-1+\eta_P} \cdot DQM_R^{-1+\eta_R} \cdot D_0 \cdot m^2 \cdot \eta_P \cdot \eta_R}{(-1 + m \cdot DQM_P^{\eta_P})^2} \\ \frac{DQM_P^{-1+\eta_P} \cdot DQM_R^{-1+\eta_R} \cdot D_0 \cdot m^2 \cdot \eta_P \cdot \eta_R}{(-1 + m \cdot DQM_P^{\eta_P})^2} & \frac{-DQM_P^{-2+\eta_P} \cdot m \cdot (\lambda^* + DQM_R^{\eta_R} \cdot D_0 \cdot m) \cdot \eta_P \cdot (-1 + \eta_P + DQM_P^{\eta_P} \cdot m \cdot (1 + \eta_P))}{(-1 + m \cdot DQM_P^{\eta_P})^3} \end{array} \right) \quad (2.3)$$

Wie leicht zu sehen ist, gilt für den ersten Eintrag der Matrix

$$\frac{\partial f}{\partial DQM_P \partial DQM_R}(DQM_P, DQM_R) < 0. \text{ Das zweite Hauptminorenkriterium ist nicht ohne weiteres}$$

zu zeigen, da zwar die Kreuzableitung $\frac{\partial f}{\partial DQM_P \partial DQM_R}(DQM_P, DQM_R) > 0$ ist, allerdings ist das

Vorzeichen des Eintrags $\frac{\partial f}{\partial DQM_P \partial DQM_P}(DQM_P, DQM_R) > 0$ von den Ausprägungen der

Variablen abhängig. Jedoch kann nach Umformung der Determinante der Hessematrix eine

Bedingung (\mathcal{A}) angegeben werden, mittels derer das zweite Hauptminorenkriterium

$H f(DQM_P, DQM_R) > 0$ erfüllt ist. Die Bedingung (\mathcal{A}) ist von folgender Form:

$$\underbrace{(\lambda_{-DQM} \cdot (-1 + \eta_P + DQM_P^{\eta_P} \cdot m \cdot (1 + \eta_P)) \cdot (-1 + \eta_R) + DQM_R^{\eta_R} \cdot D_0 \cdot x \cdot m \cdot ((-1 + \eta_P) \cdot (-1 + \eta_R) + DQM_P^{\eta_P} \cdot m \cdot (-1 - \eta_P + \eta_R)))}_{>0} \cdot \underbrace{(\alpha'')}_{>0} \cdot \underbrace{(\alpha''')}_{>0} > 0 \quad (\mathcal{A})$$

Bei der Analyse der Bedingung fällt auf, dass DQM_R nur auf den Term (α''') wirkt. Im Fall

$(\alpha''') < 0$ führt $(DQM_R \rightarrow 1)$ dazu, dass (α''') maximal negativ wird, was zugleich am

ungünstigsten für eine gültige Bedingung (\mathcal{A}) ist. Wenn für diesen Fall jedoch gezeigt werden

kann, dass für alle $DQM_P \in [0; 1]$ die Bedingung (\mathcal{A}) gilt, so gilt die Bedingung (\mathcal{A}) auch für

$(DQM_P, DQM_R) \in [0; 1] \times [0; 1]$. Damit wäre die Determinante der Hesse-Matrix

$H f(DQM_P, DQM_R) > 0$ und das zweite Hauptminorenkriterium erfüllt.

Um zu zeigen, dass die Bedingung (\mathcal{A}) für alle $DQM_P \in [0; 1]$ gilt, kann (\mathcal{A}) mit

$(DQM_R \rightarrow 1)$ umgeformt und nach DQM_P aufgelöst werden. Es ergibt sich:

$$\begin{array}{c}
\begin{array}{c}
\text{>0} \\
\frac{DQM_P^{-2+\mu_P} \cdot k_P \cdot (\lambda_{-DQM} + D_0 \cdot m \cdot DQM_R^{\eta_R}) \cdot (-DQM_P^{2\eta_P} \cdot m^2 \cdot (\eta_P - \mu_P) \cdot (1 + \eta_P - \mu_P) + \mu_P - \mu_P^2 + DQM_P^{\eta_P} \cdot m (\eta_P - \eta_P^2 - 2\eta_P \mu_P + 2(-1 + \mu_P) \cdot \mu_P))}{(-1 + m \cdot DQM_P^{\eta_P})^2} \\
\text{>0}
\end{array}
\qquad
\begin{array}{c}
\text{>0} \\
\frac{DQM_P^{-1+\eta_P} \cdot DQM_R^{-1+\eta_R} \cdot D_0 \cdot k_P \cdot m \cdot \eta_R \cdot (DQM_P^{\eta_P} \cdot m (\eta_P - \mu_P) + \mu_P)}{(-1 + m \cdot DQM_P^{\eta_P})^2} \\
\text{<0}
\end{array}
\end{array}
\left. \vphantom{\begin{array}{c} \text{>0} \\ \text{>0} \end{array}} \right\}
\begin{array}{c}
\frac{DQM_P^{-1+\eta_P} \cdot DQM_R^{-1+\eta_R} \cdot D_0 \cdot k_P \cdot m \cdot \eta_R \cdot (DQM_P^{\eta_P} \cdot m (\eta_P - \mu_P) + \mu_P)}{(-1 + m \cdot DQM_P^{\eta_P})^2} \\
\text{>0}
\end{array}
\qquad
\begin{array}{c}
\frac{DQM_P^{\eta_P} \cdot DQM_R^{-2+\eta_R} \cdot D_0 \cdot k_P \cdot m \cdot (-1 + \eta_R) \cdot \eta_R}{1 - m \cdot DQM_P^{\eta_P}} \\
\text{<0}
\end{array}
\end{array}
\quad (2.4)$$

Wie man sieht, ergibt sich für den ersten Teil der Kostenfunktion (5) für $(DQM_P, DQM_R) \in [0;1] \times [0;1]$ ein konvexer Verlauf, da der erste Eintrag der Hesse-Matrix > 0 und die Determinante der Hesse-Matrix zugleich < 0 ist. Da der zweite Teil der Kostenfunktion (DQM_R) offensichtlich konvex ist, ergibt sich bei einer additiven Verknüpfung für die gesamte Kostenfunktion ein konvexer Verlauf. Da die konvexe Kostenfunktion von der konkaven Bruttonutzenfunktion abgezogen wird, ergibt sich somit eine konkave Nettonutzenfunktion, wenn die Bedingung (\mathcal{A}') erfüllt ist.

Die Bedingung (\mathcal{A}') ist somit hinreichend, damit die in den Termen (6) und (7) beschriebenen Extremstellen für DQM_R und DQM_P globale Maxima darstellen.

Anhang 3: Ausführungen zum Fallbeispiel des FDL

Im Anhang soll zur Ergänzung der obigen Ausführungen die konkrete Ausgestaltung der DQM erläutert werden.

Bei den reaktiven DQM könnte der FDL potenziell eine Reihe von Algorithmen zur Suche nach Assoziationsregeln (hier: Auffinden aller Ereigniskombinationen, wie zum Beispiel $\{z1, z6, z9\}$ für $z1$: *Kunde hat zwei Kinder*, $z6$: *Kunde ist zu mehr als 50% in Aktien investiert* und $z9$: *Kunde ist Beamter* und deren Häufigkeiten) einsetzen. Um diese Suche effizient zu gestalten, müssen Äquivalenzklassen gebildet werden, das heißt zwei n -Ereigniskombinationen, wobei n die Anzahl der Ereignisse mit $n > 1$ angibt, gehören zu derselben Äquivalenzklasse, wenn sie in den ersten $(n-1)$ -Ereignissen übereinstimmen Präfix haben. Die 3-Ereigniskombinationen $\{z1, z6, z9\}$ und $\{z1, z6, z7\}$ gehören demzufolge eindeutig zur Äquivalenzklasse $E(\{z1, z6\})$. Wird der Suchraum anhand der Äquivalenzklassen vollständig dargestellt, lässt sich auf der Menge aller Äquivalenzklassen ein Baum definieren (Klasse $E(\{z1, z6\})$ ist ein Sohn von Klasse $E(\{z1\})$). Dieser Baum und damit der Suchraum kann nunmehr bei Angabe einer Häufigkeitsgrenze (Höhe des Mindestsupports beziehungsweise der -konfidenz) effizient beschnitten werden, da sobald eine Äquivalenzklasse diese Grenze unterschreitet, alle Äquivalenzklassen unterhalb dieser Äquivalenzklasse ebenfalls keine häufigen Ereigniskombinationen mehr enthalten können. Für die Suche entlang der Pfade des aufgrund der Vielzahl von Ereignissen großen Baums könnten vom FDL generell Algorithmen Verwendung finden, die eine Breiten- oder eine Tiefensuche verfolgen. Aus Gründen der Datenbankimplementierung und aus toolorientierten Gründen soll beziehungsweise muss konkret auf Algorithmen zurückgegriffen werden, die eine Tiefensuche zur Generierung von Assoziationsregeln durchführen. Tiefensuche bedeutet, dass die Pfade, ausgehend von der Baumwurzel, zu den einzelnen Blättern (rekursiv) verfolgt werden.

Ein Algorithmus, der die Tiefensuche verwendet, ist der Frequent-Pattern (FP)-Growth Algorithmus (vergleiche *Han et al. (2000)*). Da er auf einer als Frequent-Pattern-Baum aufbereiteten Datenbasis operiert, ist er aus technischer Sicht gegenüber vergleichbaren Algorithmen effizient. Deswegen wurde er hier zugrunde gelegt. Basis der Tiefensuche bildet dabei die Menge aller häufigen Ereignisse als häufige 1-Ereigniskombinationen. In absteigender Reihenfolge (im Sinne der Relation $<$) wird ausgehend von jedem dieser Ereignisse rekursiv im Suchraum abgestiegen, indem zu der jeweils bearbeiteten Ereigniskombination ein einzelnes Ereignis hinzugenommen wird, welches zusammen mit dieser in mindestens einer Transaktion enthalten ist. Um Ereigniskombinationen nicht mehrfach zu zählen, werden diese nur um solche Ereignisse erweitert, die kleiner (im Sinne der Relation $<$) sind als sämtliche in der Ereigniskombination bereits enthaltenen Ereignisse. Der Suchraum wird folglich strikt von rechts nach links und von oben nach unten durchlaufen. Dabei wird nur zu tatsächlich in den Transaktionen vorkommenden Ereigniskombinationen abgestiegen.

Nach der Auswahl des (FP)-Growth Algorithmus im ersten Schritt, stellt sich die Frage nach dessen Konfiguration anhand der Häufigkeitsgrenze für die Baumsuche. Dies ist deshalb wichtig, um die Anzahl der generierten und für die zur Umsetzung zu prüfenden Assoziationsregeln (ist eine erhaltene DQ-Regel inhaltlich wirklich sinnvoll und sollte deshalb realisiert werden?) zu steuern. Wird die Häufigkeitsgrenze dabei auf eins gesetzt, wird die maximale Anzahl an Assoziationsregeln generiert.

Die Entscheidung über die Häufigkeitsgrenze und damit den Mindestsupport beziehungsweise die -konfidenz kann anhand des optimalen Umfangs an DQM_R (im Fallbeispiel ist $DQM_{R,opt} = 0,5$) unterstützt werden. So sind demzufolge nicht, wie bei einer größeren Datenmenge sonst in der Praxis üblich, umfangreiche reaktive DQM zu ergreifen und damit eine hohe Anzahl von Assoziationsregeln zu ermitteln. Vielmehr sind Mindestsupport und -konfidenz so zu wählen, dass nur circa 50% der insgesamt ermittelbaren Regeln selektiert werden. Dies ist im Beispiel bei einer mit Wahrscheinlichkeit von 75% (Mindestkonfidenz) und einer absoluten Auftrittshäufigkeit von 8mal bei 3-Ereigniskombinationen gegeben. Derartige Vorgaben sind deshalb wichtig, da DQ-Tools (zum Beispiel WizRule® von WizSoft) dem Anwender völlig freistellen, wie die Konfigurationsparameter (beispielsweise Minimum probability of if-then-rules) zur Bestimmung und Auswahl der Assoziationsregeln gesetzt werden. Die Vorgaben bieten somit eine wichtige Unterstützung (neben anderen Kriterien wie Ressourcen- oder Zeitrestriktionen), um den gewünschten Regelumfang für die zugrunde gelegte Datenmenge zu gewinnen und um keine Über- oder Unterdimensionierung (beispielsweise als Ergebnis erhält man nur tautologische Regeln) des Regelumfangs zu erhalten.

Bei den proaktiven DQM werden bei der Erfassung von Daten standardisierte Begriffe vorgegeben. Dies ermöglicht eine von der eingebenden Person weitgehend unabhängige Dokumentation. Viele Fehler, die gerade im praktischen Umfeld auf eine inkonsistente Begriffsverwendung zurückgehen, können somit vermieden werden. Für die Interpretierbarkeit und richtige Auswertung der Daten spielt dies im oben vorgestellten Ansatz beispielsweise bei der Individualisierung (Ermittlung der Kundencharakteristika) oder bei Produktkampagnen eine entscheidende Rolle.

Konkret bedeutet dies für die Beratungsapplikation, dass für einzelne Eingabefelder mittels einer Drop-Down-Box eine Auswahl erfolgen kann, wobei die Bedeutung der angebotenen Eingaben bei Bedarf durch den Berater abzufragen ist. Wird durch den Berater kein vorgegebener Wert selektiert, sondern eine freie Eingabe durchgeführt, so wird analysiert, ob der Begriff bereits in einem anderen Zusammenhang (beispielsweise in einem anderen Eingabefeld) verwendet wird. Ist dies der Fall, wird der Berater auf die im System hinterlegte Begriffsbedeutung hingewiesen und kann diese nochmals ändern. Potenzielle Inkonsistenzen können so frühzeitiger erkannt und vermieden werden. Daneben werden die einzelnen Eingaben

des Beraters, selbst wenn er jeweils eine Auswahl aus der Drop-Down-Box vornimmt, auf wechselseitige Konsistenz geprüft. Einfache Beispiele sind die unmittelbare Kontrolle der Konsistenz zwischen Postleitzahl und Ort nach Eingabe der Kundenanschrift oder der Zusammenhang zwischen Steuerklasse und dem Familienstand. Anspruchsvollere Überprüfungen beziehen sich beispielsweise auf potenzielle Widersprüche im Anlageverhalten des Kunden im Zeitverlauf (Initiierung durch die Eingabe einer neuen Transaktion), die auf (nicht erfasste) Veränderungen in den Stammdaten des Kunden oder seiner Charakteristika beruhen. Auf die Widersprüche wird der Berater hingewiesen und kann im Bedarfsfall eine Nacherfassung oder Korrektur der Daten vornehmen.

Der Umfang an proaktiven DQM spiegelt sich dabei im Umfang der Begriffsstandardisierung und der damit einhergehenden Konsistenzchecks wider. Ein Wert von $DQM_{p,opt} = 0,6$ ist dabei so zu interpretieren, dass circa 60% der meistbenutzten Begriffe im System (zum Beispiel dokumentierte Kundenäußerungen und -meinungen bezüglich Aktien) standardisiert und bei der Erfassung geprüft werden. Vernachlässigt werden darf dabei nicht der Zeitaufwand, den ein Berater nach einem Systemhinweis auf eine Inkonsistenz benötigt, diese zu prüfen und zu korrigieren. Jedoch führt diese Zeitinvestition zu einer direkten Rückkoppelung mit dem Kunden und einer gesicherten Korrektur im Gegensatz zu statistischen Zusammenhängen (wie bei den Assoziationsregeln), deren empirische Korrektheit (versus Konsistenz bei Assoziationsregeln) immer noch gesondert überprüft werden muss.

Anhang 4: Konkavitätseigenschaften der Nettonutzenfunktion, Randlösungen im Fallbeispiel und Ausführungen zum Signifikanzniveau

Um unter mathematischen Gesichtspunkten zu zeigen, dass die im Beispiel des Kapitels 4 ermittelten Werte für $DQM_{P,opt}$ und $DQM_{R,opt}$ ein globales Maximum im Definitionsbereich ergeben, ist die Konkavitätseigenschaften der Nettonutzenfunktion und somit die hinreichende Bedingung (\mathcal{A}') zu prüfen. Es ergibt sich:

$$DQM_P < \left[\frac{-(0.33 + 0.42 \cdot 0.1) \cdot (-1 + 0.34) \cdot (-1 + 0.25)}{0.1 \cdot (0.33 \cdot (1 + 0.34) \cdot (-1 + 0.25) + 0.42 \cdot 0.1 \cdot (-1 - 0.34 + 0.25))} \right] \frac{1}{0.34}$$

$$\Leftrightarrow DQM_P < 105.79 \quad (\mathcal{A}')$$

Die Bedingung (\mathcal{A}') ist im Beispiel für $DQM_{P,opt} \in [0; 1]$ erfüllt, das heißt die Nettonutzenfunktion ist im gesamten Definitionsbereich konkav.

Zudem sind ebenso die beiden Randlösungen zu untersuchen, bei welchen die fixen Kostenbestandteile entfallen. Hier ergibt sich:

Proaktive Maßnahmenintensität DQM_P	Reaktive Maßnahmenintensität DQM_R	Zus. Transaktionsanteil λ_{DQM}	Zus. Bruttonutzen $u \cdot \lambda_{DQM}$ (in Mio. €)	Proaktive Qualitätskosten $K_P + k_P \cdot DQM_P^{\mu_P} \cdot D_I$ (in Mio. €)	Reaktive Qualitätskosten $K_R + k_R \cdot DQM_R^{\mu_R} \cdot D_O$ (in Mio. €)	Nettonutzen E (in Mio. €)
0	0,47	0,035	0,67	0	0,277	0,393
0,6	0	0,03	0,575	0,15	0	0,425

Die Tabelle legt dar, dass im Beispiel die jeweiligen Nettonutzen der Randlösungen geringer sind als der Nettonutzen für die optimalen Maßnahmenintensitäten.

Am Ende des Kapitels 3 wurde das Vorgehen bei Existenz eines Signifikanzniveaus λ_{sig} ausgeführt. Grundsätzlich gilt hier, dass wegen der Nettonutzen ausgehend vom Optimum bei einer weiteren Steigerung von DQM monoton abnimmt. Wenn diese Steigerung jedoch notwendig ist, um das Signifikanzniveau zu überwinden, ist es zweckmäßig denjenigen Umfang an DQM_P und DQM_R zu prüfen, mittels dessen gerade das Signifikanzniveau und damit das korrespondierende λ_{sig} erreicht werden kann. Eine weitere Erhöhung von DQM_P beziehungsweise DQM_R vermindert den Nettonutzen. Die effiziente Kombination von DQM_P und DQM_R zur Erreichung von λ_{sig} ist dabei unter Berücksichtigung der proaktiven und reaktiven Wirkungs- und Kostenfaktoren zu ermitteln. Liegt im Fallbeispiel das Signifikanzniveau bei $\lambda_{sig} = 0,41$ und ist somit größer als $\lambda_{-DQM} + \lambda_{DQM} = 0,33 + 0,069 = 0,399$, so wäre der optimale Einsatz an DQM zwecklos. Insofern müsste mindestens ein DQ-induzierter Transaktionsanteil von $\lambda_{sig} - \lambda_{-DQM} = \lambda_{DQM} = 0,08$ erzielt werden, bei dem zugleich unter

Berücksichtigung des Signifikanzniveaus der maximale Nettonutzen resultiert. Setzt man $\lambda_{DQM} = 0,08$ und erhält damit alle entsprechenden (DQM_P, DQM_R) -Kombinationen, die zu diesem DQ-induzierten Transaktionsanteil führen, so lässt sich mit Hilfe der Nettonutzenfunktionen diejenige Kombination ermitteln, mit der der maximale Nettonutzen realisierbar ist. Im Beispiel ist dies die Kombination $(DQM_P, DQM_R) \rightarrow (0,951, 0,797)$, die zu einem Nettonutzen E in Höhe von 0,61 Mio. € führt. Im Vergleich dazu ergab sich der Nettonutzen E zuvor bei $\lambda_{DQM} = 0,069$ zu 0,677 Mio. €.