

Modellgestützte Nachfrageanalyse im Internet als Grundlage für eine kundenorientierte Angebotspolitik

Reinhold Decker and Sören W. Scholz

Department of Business Administration and Economics
Bielefeld University – Germany

Address correspondence to:

Prof. Dr. Reinhold Decker
Bielefeld University
Department of Business Administration and Economics
PO Box 10 01 31
D-33501 Bielefeld
Germany

Phone: +49 (0)521-1066913
Fax: +49 (0)521-1066456
E-mail: rdecker@wiwi.uni-bielefeld.de
sscholz@wiwi.uni-bielefeld.de

Modellgestützte Nachfrageanalyse im Internet als Grundlage für eine kundenorientierte Angebotspolitik

Zusammenfassung

Im vorliegenden Beitrag wird ein neues Modell auf Basis der Poisson-Regression zur Analyse der Nachfrage nach aktuellen Internetangeboten vorgestellt. Die gewählte Vorgehensweise weist Analogien zur Conjoint-Analyse auf und ermöglicht Einblicke in die der Inanspruchnahme bzw. Nichtinanspruchnahme alternativer Leistungsangebote zugrunde liegenden Präferenzstrukturen. Im Gegensatz zur Conjoint-Analyse wird dabei auf Dialogdaten zurückgegriffen, wie sie im Zusammenhang mit individuellen Leistungsanfragen im Internet automatisch anfallen. Das Modell liefert auf diese Weise Ansatzpunkte zur Verbesserung des gegenwärtigen Angebotsspektrums. Die prinzipielle Vorgehensweise wird am Beispiel von Nutzungsdaten eines kommerziellen Internetportals und Online-Buchungsdaten eines bekannten Reiseveranstalters demonstriert.

Schlüsselwörter: Analyse von Internetangeboten, Log-File-Daten, Neuproduktentwicklung, Poisson-Regression, Responsemodellierung

Abstract

This paper presents a new model based on Poisson regression for analyzing consumer demands regarding current internet offerings. Our approach bears resemblance to conjoint analysis and delivers insights into the preference structures underlying the usage or non-usage of alternative service offerings. In contrast to conjoint analysis we use transaction data as they automatically result when a service request is launched. In doing so, the model provides useful information with respect to potential improvements of the current range of products. The basic idea of our approach is demonstrated by means of usage data provided by a commercial internet portal and online booking data provided by a well-known tour operator.

Keywords: Analysis of internet offerings, log file data, new product development, Poisson regression, response modelling

1. Motivation und Literatur

Die Analyse des Kauf- und Konsumverhaltens im Internet ist heute eine zentrale Aufgabe der Marketingforschung (Cheung et al. (2003)). Entsprechend existiert mittlerweile eine unüberschaubare Menge von Beiträgen, die sich mit den Vor- und Nachteilen des Internets als Basis für den Kauf oder Verkauf von Leistungen unterschiedlichster Art auseinandersetzen (Steckel et al. (2005)). In einzelnen Bereichen, so z. B. im Tourismus oder in der Musikbranche, erfolgt bereits heute ein beträchtlicher Teil der Kaufhandlungen unter Einbeziehung des Internets. Eine anspruchsvolle Aufgabe stellt in diesem Zusammenhang die Identifikation von Verhaltensmustern dar, mit deren Hilfe der Frage nachgegangen werden kann, wann und warum es bei einem Internetnutzer zu einer Kaufhandlung kommt (Rudolph et al. (2002), Bauer et al. (2004), Constantinides (2004), Montgomery et al. (2004)).

Ein möglicher Ansatz, um solche Verhaltensmuster zu ergründen, ist die Durchführung von Primärstudien, zumeist Befragungen von Internetnutzern. Erfolgen diese online, so erweist sich die hohe Flexibilität des Mediums Internet bei der Ausgestaltung der Frageinhalte als vorteilhaft, z. B. die Filterführung betreffend (Meißner et al. (2008), Dahan und Srinivasan (2000)). Als Nachteil muss hingegen die Tatsache gewertet werden, dass die Befragten, je nach Thema, zu sozial erwünschten Antworten neigen können. Alternativ besteht die Möglichkeit einer modellgestützten Analyse der bei realen Kundenkontakten im Internet aufgezeichneten Daten, wie sie auch beim Web Usage Mining Verwendung finden (Spiliopoulou (2000)). Der vorliegende Beitrag verfolgt diesen zweiten Ansatz und kommt somit ohne eigene Datenerhebung aus. Die Zweckmäßigkeit der sekundädatengestützten Methodik wird anhand zweier realer Beispiele demonstriert.

Das erste Beispiel bezieht sich auf Internetportale. Ein Internetportal bietet seinen Nutzern Leistungen, die sich, wie das in Abbildung 1 dargestellte Beispiel dokumentiert, aus unterschiedlichen Bestandteilen (Rubriken) zusammensetzt. Im vorliegenden Fall reicht das Spektrum von der Vermittlung von Immobilien bis hin zur Bereitstellung aktueller Börsendaten. Die unter der Überschrift „Top-Themen“ zusammengefassten Rubriken stellen im Sinne der nachfolgend beschriebenen Methodik Leistungsmerkmale dar, die ihrerseits wiederum in verschiedenen Ausprägungen vorliegen können und zumeist nominales Skalenniveau aufweisen. Das Merkmal „Auto“ beispielsweise besaß zum Zeitpunkt der Erstellung dieses Beitrags u. a. die thematisch sortierten Ausprägungen (Unterrubriken) „Gebrauchtwagen“, „Neuwagen“, „Routenplaner“, „Rallye“, „Finanzierung“ und „Autogalerien“.



Abb. 1: Möglicher Aufbau der Merkmalsebene eines Internetportals am Beispiel von Lycos

Neben den unterschiedlichen Rubriken enthalten die einzelnen Seiten eines Internetportals in der Regel auch Werbeanzeigen (z. B. in Form von Banner oder laufenden Textbändern), deren Wirkung umso stärker zum Tragen kommen kann, je länger sich die Besucher des Portals auf der betreffenden Web-Seite aufhalten. Die Aufenthaltsdauer auf den Seiten eines Internetportals (z. B. gemessen in Minuten) kann somit als Maß für seine Attraktivität aus Nutzersicht interpretiert werden und stellt im Folgenden die metrisch skalierte Responsevariable dar. Diese Vorgehensweise ist in der Web Usage-Analyse weit verbreitet (siehe z. B. Burklen et al. (2005)). Tao et al. (2006) diskutieren in diesem Zusammenhang die Problematik, dass in Log-Files im Normalfall nur der Aufruf und das Verlassen der Seite gemessen werden. Eine genauere Messung der Nutzungsdauer einzelner Seiten erlaubt der Einsatz von Cookies oder 1-Pixel-Gifs, die in sehr kurzen Abständen auf der aufgerufenen Seite ausgetauscht werden.

Die Attraktivität eines Internetportals resultiert im Wesentlichen aus dem Nutzen der angebotenen Leistungsmerkmale resp. Rubriken. Gleichzeitig wird eine Web-Seite für auf diesem Wege Werbung betreibende Unternehmen umso interessanter, je höher die Aufenthaltsdauer potenzieller Kunden auf der betreffenden Seite ausfällt, womit sich die Kausalkette schließt. Dies führt schließlich zu der Frage, welchen Beitrag die einzelnen Leistungsmerkmale zur Gesamtattraktivität des Portals liefern. Zuverlässige Erkenntnisse hierüber können z. B. dazu beitragen, bestehende Portale noch kundengerechter zu gestalten. Als Analysegrundlage bieten sich die in Log-Files aufgezeichneten Portalnutzungsdaten an. Welche technischen Probleme sich im Zusammenhang mit der analysegerechten Aufbereitung von Log-File-Daten ergeben können, wird z. B. bei Cooley et al. (1999) thematisiert und ist nicht Gegenstand der weiteren Ausführungen.

Das zweite Beispiel entstammt dem Bereich Tourismus. Mittlerweile haben sich bereits 39 Prozent der Deutschen schon einmal online zu Urlaubsfragen informiert; 19 Prozent haben sogar schon einmal eine Urlaubsreise online gebucht (FUR (2007)). Nahezu alle Touristikunternehmen bieten ihre Leistungen heute auch oder sogar ausschließlich im Internet an. Der Kunde hat hier die Möglichkeit, sich eigenständig über alternative Reiseangebote zu informieren und auch zu buchen (Kuom und Oertel (1999)). Der anhaltende Trend zu Online-Reisebuchungen wird vor allem durch schnellere und einfachere Reservierungs- und Buchungsmöglichkeiten, die Bündelung von aktuellen Informationen und Angeboten zu einem Reisegebiet, und nicht zuletzt die Reduktion der Kosten durch effizientere interne Operationen (z. B. dem Umgehen von Handelsmittlern in Form von konventionellen Reisebüros) verstärkt (Oertel et al. (2003)). Ausgangspunkt der nachfolgenden Überlegungen sind Transaktionsdaten, wie sie z. B. bei einem Dialog der durch Abbildung 2 angedeuteten Art entstehen können.

BUY.by PAUSCHAL
Hotline: 0211-1770346

REISEDATEN

1 Flugauswahl 2 Hotelauswahl 3 Zusätze & Buchung 4 Bestätigung

von: Düsseldorf
nach: Palma de Mallorca

Hinflug: 31 Juli 07
Rückflug: 7 August 07
Klasse: Economy

Erwachsene: 2
Kinder: 0 (2-11 Jahre)
Babies: 0 (0-23 Monate)

Flüge

1
↓ **Flugauswahl**
Wählen Sie Ihr gewünschtes **Flugziel**, eine Flugklasse und Ihre Reisedaten.

2
↓ **Hotelauswahl**
Wählen Sie ein passendes **Hotel** oder **Appartement**.

3
↓ **Zusätze & Buchung**
Buchen Sie abschließend opt. einen **Transfer / Mietwagen, Rail & Fly** und einen **Reiseschutz**.

Abb. 2: Elemente einer Leistungsspezifikation am Beispiel von BUY.by

Hierbei stellen beispielsweise der Abflugort und das Reiseziel Merkmale der Leistung „vermittelte Reise“ dar. Mögliche Ausprägungen der genannten Merkmale sind z. B. Düsseldorf bzw. Palma de Mallorca. In beiden Fällen handelt es sich um nominale Merkmale, während hingegen das aus Hin- und Rückflugdatum resultierende Merkmal Reisedauer (in Tagen) als metrisch zu interpretieren ist. Weitere Merkmale können aus der Hotelauswahl sowie zusätzlich angebotenen Services wie Mietwagen und Reiseschutz resultieren. Besucht eine Person

die betreffenden Web-Seiten eines Reiseveranstalters, so kann dies die folgenden Gründe haben:

1. Sie hat die konkrete Absicht, auf diesem Wege eine Reise zu buchen („Buchung“).
2. Sie möchte sich ohne konkrete Buchungsabsicht über die vom kontaktierten Reiseveranstalter angebotenen touristischen Offerten informieren („Anfrage“).
3. Sie landet aus nicht mit einem konkreten Reiseinteresse verbundenen Gründen auf den Web-Seiten.

Der zuletzt genannte Fall kann bei den weiteren Betrachtungen außer Acht bleiben. Die beiden anderen Punkte werfen hingegen die unter Produktgestaltungsgesichtspunkten interessante Frage auf, welche im Dialog spezifizierten Kombinationen von Merkmalsausprägungen häufig zu einer Online-Buchung führen und welche eher in einer der Informationsbedarfsbefriedigung dienenden Anfrage ihren Niederschlag finden (Brockhoff (1999)).

Die Beantwortung der anhand der beiden Beispiele exemplarisch aufgeworfenen Fragen liefert nicht nur Hinweise auf die Attraktivität der gegenwärtig angebotenen Leistungen, sondern sie kann, was unter Planungsgesichtspunkten häufig noch erheblich interessanter ist, auch Anhaltspunkte für die Ausgestaltung der zukünftig bereitzustellenden Leistungen geben. Insofern weist die in Abschnitt 2 vorgestellte Methodik Analogien zum Ansatz der Conjoint-Analyse auf (siehe z. B. Backhaus et al. (2007), Skiera und Gensler (2002)), kommt aber im Gegensatz zu dieser ohne eine eigens durchzuführende und damit zumeist kostspielige Datenerhebung aus und ist nicht-linear in den Parametern. Bei Conjoint-Analysen stellt darüber hinaus die Anzahl der einbezogenen Merkmale und Merkmalsausprägungen eine immer wieder thematisierte Problematik dar (Meißner et al. (2007)).

2. Responsemodellierung mittels Poisson-Regression

2.1 Annahmen und Datenstruktur

Die nachfolgenden Modellierungsüberlegungen basieren auf drei zentralen Annahmen:

- 1) Die zu analysierenden Leistungen stellen Kombinationen von Ausprägungen einer festen Anzahl von Leistungsmerkmalen dar.
- 2a) Die Aufenthaltsdauer in einem Internetportal kann als Maß für dessen Attraktivität interpretiert werden.
- 2b) Das Zustandekommen einer Online-Buchung resp. der damit verbundene Kaufakt dokumentiert die grundsätzliche Akzeptanz der betreffenden Leistung.

Um die Struktur der für die Analyse heranzuziehenden Daten beschreiben zu können, werden einige Bezeichner benötigt. Es sei

$k = 1, \dots, K$ der Index für die individuellen Leistungsspezifikationen,

$l = 1, \dots, L$ der Index für die Leistungsmerkmale und

$m_l = 1, \dots, M_l$ der Index für die möglichen Ausprägungen von Merkmal l .

Unter Verwendung der damit definierbaren Variablen

$$x_{klm_l} = \begin{cases} 1, & \text{falls bei Spezifikation } k \text{ das Merkmal } l \text{ in Ausprägung } m_l \text{ gewählt wurde} \\ 0, & \text{sonst} \end{cases},$$

$$y_{k1} = \begin{cases} 1, & \text{falls die Spezifikation } k \text{ lediglich einer Anfrage diene} \\ 0, & \text{sonst} \end{cases},$$

$$y_{k2} = \begin{cases} 1, & \text{falls bei Spezifikation } k \text{ eine Buchung zustande kam} \\ 0, & \text{sonst} \end{cases}$$

und $y_{k3} \in \{0, 1, 2, \dots\}$ als der Dauer einer Spezifikation bzw. Konsultation k des betrachteten Portals in Minuten ergibt sich die durch Tabelle 1 veranschaulichte Datengrundlage. Dabei beschreiben die Kombinationen der L Leistungsmerkmale mit den Responsevariablen Anfragestatus und Buchungstatus die durch das Tourismusbeispiel veranschaulichte Datengrundlage (vgl. Abbildung 2). Hiernach wurde beispielsweise bei Spezifikation 2 das Merkmal 1 in Ausprägung 1, das Merkmal L hingegen in der Ausprägung 3 gewählt. Die so spezifizierte Leistung führte zu einer Buchung ($y_{k2} = 1$). Dabei muss allgemein gelten: $\sum_{m_l} x_{klm_l} = 1 \forall k, l$ und $y_{k2} + y_{k1} = 1 \forall k$. Jedes der als nominal skaliert angenommenen Leistungsmerkmale muss also in genau einer Ausprägung vorliegen. Die zweite Bedingung bringt zum Ausdruck, dass eine Leistungsspezifikation entweder der Information („Anfrage“) oder dem Vollzug eines Kaufakts („Buchung“) dient. Der Fall $y_{k2} + y_{k1} = 0$, z. B. infolge des vorzeitigen Abbruchs einer Spezifikation, liefert für die vorliegende Fragestellung keine brauchbaren Erkenntnisse und bleibt deshalb von den weiteren Betrachtungen ausgeschlossen.

Spezifikation k	Merkmal 1 $x_{k11}, \dots, x_{k1M_1}$...	Merkmal L $x_{kL1}, \dots, x_{kLM_L}$	Anfrage- status y_{k1}	Buchungs- status y_{k2}	Konsultations- dauer y_{k3}
1	0,0,0,...,0,1	...	0,1,0,...,0,0	1	0	3
2	1,0,0,...,0,0	...	0,0,1,...,0,0	0	1	12
\vdots	\vdots		\vdots	\vdots	\vdots	\vdots
K	0,0,0,...,0,1	...	0,1,0,...,0,0	0	0	1

Tab. 1: Struktur der disaggregierten Beobachtungsdaten

In analoger Weise spezifiziert eine Kombination der Leistungsmerkmale mit der Responsevariable Konsultationsdauer die durch das Portalbeispiel veranschaulichte Datengrundlage (vgl. Abbildung 1). In diesem Fall ist es jedoch durchaus möglich, dass mehrere Ausprägungen eines Merkmals gleichzeitig den Wert 1 annehmen. Besitzen alle Ausprägungen aller Merkmale den Wert 0, so bedeutet dies, dass bei der betreffenden Konsultation zwar auf die Eingangsseite des Portals zugegriffen, auf einen weiteren Einstieg in das Portal aber verzichtet wurde. Die betreffenden Datensätze bleiben bei der Analyse unberücksichtigt. Die Spezifikation 1 in Tabelle 1 wäre somit wie folgt zu interpretieren: Im Rahmen dieser Konsultation wurde in Rubrik 1 auf die Unterrubrik M_1 und in Rubrik L auf die Unterrubrik 2 zugegriffen. Die Zugriffe führten in Summe zu einer Gesamtaufenthaltsdauer im Portal von drei Minuten.

Die durch Tabelle 1 veranschaulichten Rohdaten können nun in der Weise aggregiert werden, dass im Falle des Tourismusbeispiels die Spezifikationen identischer Leistungen zusammengefasst und mit den entsprechenden Häufigkeiten für erfolgte Anfragen und Buchungen versehen werden. Mit $I = \prod_{l=1}^L M_l$ als der Anzahl möglicher Kombinationen von Merkmalsausprägungen resp. Leistungen und z_{i1} bzw. z_{i2} als die auf Spezifikation i entfallende Anzahl an Anfragen bzw. Buchungen ergibt sich die in Tabelle 2 dargestellte Datenstruktur. Auf Spezifikation 2 entfallen demzufolge 34 Anfragen und 5 Buchungen.

Spezifikation i	Merkmal 1 $x_{i11}, \dots, x_{i1M_1}$...	Merkmal L $x_{iL1}, \dots, x_{iLM_L}$	Anzahl Anfragen z_{i1}	Anzahl Buchungen z_{i2}
1	1,0,0,...,0,0	...	1,0,0,...,0,0	12	2
2	1,0,0,...,0,0	...	0,1,0,...,0,0	34	5
\vdots	\vdots		\vdots	\vdots	\vdots
I	0,0,0,...,0,1	...	0,0,0,...,0,1	4	0

Tab. 2: Struktur der aggregierten Touristikdaten

Im Falle der Auswertung von Portalnutzungsdaten ist eine solche Aggregation nicht erforderlich. Hier liegt mit Responsevariable y_{k3} bereits die für die modellgestützte Analyse erforderliche Datenstruktur vor.

2.2 Analyse der Portalnutzung

Interpretiert man die Dauer des Aufenthalts im Internetportal in Minuten als Ergebnis eines Zählprozesses, so bietet sich für die Modellierung der durch die Zufallsvariable Y_3 , mit Aus-

prägung y_3 , repräsentierten Zähldaten die Poisson-Regression an (vgl. hierzu auch Maddala (1983) sowie Drèze und Zufryden (1997, 1998)). Ausgangspunkt ist hierbei die Wahrscheinlichkeitsfunktion

$$P(Y_3 = y_3) = \frac{\lambda^{y_3}}{y_3!} \exp(-\lambda)$$

der Poisson-Verteilung mit $y_3 \in \{0, 1, 2, \dots\}$ und λ als dem Erwartungswert („mittlere Konsultationsdauer“) der Zufallsvariable Y_3 .

Unter Rückgriff auf die in Tabelle 1 skizzierte Datengrundlage und unter Verwendung der Parameter β_{lm_l} ($l = 1, \dots, L$; $m_l = 1, \dots, M_l$) zur Erfassung der Stärke des Einflusses von Ausprägung m_l des Merkmals l auf die Dauer einer Konsultation des Portals lässt sich folgende einfache Reparametrisierung vornehmen:

$$\lambda = \exp\left(\beta_0 + \sum_{l=1}^L \sum_{m_l=1}^{M_l} \beta_{lm_l} x_{lm_l}\right)$$

Der Parameter β_0 kann dabei als der aus der Existenz des Portals resultierende Grundnutzen interpretiert werden. Die exponentielle Modellformulierung trägt der Nichtnegativitätsforderung von λ Rechnung.

Zur Schätzung der Parameter β_0 und $\beta_{11}, \dots, \beta_{LM_L}$ des resultierenden univariaten Poisson-Regressionmodells kann die Maximum-Likelihood-Methode herangezogen werden (vgl. hierzu auch Tutz (2000, S. 249 ff.)). Die auf Basis der K Beobachtungen (Spezifikationen) zu maximierende Log-Likelihood-Funktion lautet:

$$\begin{aligned} LL(\beta_0, \beta_{11}, \dots, \beta_{LM_L}) &= \ln\left(\prod_{k=1}^K \frac{\lambda^{y_{k3}}}{y_{k3}!} \exp(-\lambda)\right) = \sum_{k=1}^K y_{k3} \ln(\lambda) - \lambda - \ln(y_{k3}!) \\ &= \sum_{k=1}^K y_{k3} \left(\beta_0 + \sum_{l=1}^L \sum_{m_l=1}^{M_l} \beta_{lm_l} x_{klm_l}\right) - \exp\left(\beta_0 + \sum_{l=1}^L \sum_{m_l=1}^{M_l} \beta_{lm_l} x_{klm_l}\right) - \ln(y_{k3}!) \end{aligned}$$

Aufgrund des dichotomen Charakters der Leistungsmerkmale x_{klm_l} und mit Bezug auf die in Unterabschnitt 2.1 dargelegte Interpretation der Responsevariable y_{k3} können die resultierenden Schätzer $\hat{\beta}_{lm_l}$ als mit den Teilnutzenwerten aus der Conjoint-Analyse vergleichbare Größen aufgefasst werden. Der Wert des Schätzers $\hat{\beta}_{21}$ entspricht demzufolge beispielsweise dem Beitrag, den die Ausprägung 1 des Merkmals 2 zu dem über die Konsultationsdauer operationalisierten Gesamtnutzen des Portals leistet.

2.3 Analyse des Buchungsverhaltens

Analog zu den Ausführungen in Unterabschnitt 2.2 und mit Bezugnahme auf die durch Tabelle 2 veranschaulichte Datengrundlage lassen sich auch im Falle des Tourismusbeispiels die Abhängigkeitsbeziehungen zwischen den betrachteten Leistungsmerkmalen und den Responsevariablen Anfrage (Zufallsvariable Z_1 mit Ausprägung z_1) und Buchung (Zufallsvariable Z_2 mit Ausprägung z_2) unter Rückgriff auf eine Poisson-Regression modellieren. Um dabei explizit zum Ausdruck zu bringen, dass eine Spezifikation entweder in der Alternative Anfrage oder in der Alternative Buchung mündet, bietet sich die kombinierte Erfassung beider Aspekte mittels bivariater Poisson-Regression an. Die entsprechende Grundgleichung der Wahrscheinlichkeitsfunktion lautet

$$P(Z_1 = z_1, Z_2 = z_2) = \prod_{j=1}^2 \frac{\lambda_j^{z_j}}{z_j!} \exp(-\lambda_j) \quad \text{für } z_j \in \{0, 1, 2, \dots\} \text{ und } j \in \{1, 2\}$$

und liefert die Wahrscheinlichkeit dafür, dass eine bestimmte Spezifikation zu genau z_1 Anfragen und z_2 Buchungen führt. Die gesuchten Modellparameter erhält man unter Rückgriff auf I Beobachtungen aus der zu maximierenden Log-Likelihood-Funktion:

$$\begin{aligned} & LL(\beta_1, \beta_2, \beta_{111}, \dots, \beta_{2LM_L}) \\ &= \sum_{i=1}^I \sum_{j=1}^2 z_{ij} \left(\beta_j + \sum_{l=1}^L \sum_{m_l=1}^{M_l} \beta_{jlm_l} x_{ilm_l} \right) - \exp \left(\beta_j + \sum_{l=1}^L \sum_{m_l=1}^{M_l} \beta_{jlm_l} x_{ilm_l} \right) - \ln(z_{ij}!) \end{aligned}$$

Analog zum univariaten Modell wird mit dem Wert der Schätzer $\hat{\beta}_{1m_l}$ und $\hat{\beta}_{2m_l}$ der Einfluss der m_l -ten Ausprägung des Merkmals l auf die über die Anzahl der Anfragen ($j=1$) bzw. die Anzahl der Buchungen ($j=2$) operationalisierte Gesamtattraktivität der betreffenden Leistung zum Ausdruck gebracht. Die bivariate Modellierung ermöglicht, im Gegensatz zu der bei Drèze und Zufryden (1998) beschriebenen Vorgehensweise, in intuitiver Weise die simultane Optimierung des Modells hinsichtlich beider hier interessierender Kundenreaktionen.

3. Beispielanwendungen

3.1 Analyse synthetischer Touristikdaten

Der in Tabelle 3 dargestellte synthetische Datensatz mit zwei Leistungsmerkmalen ist an das Tourismusbeispiel angelehnt und zeigt auf anschauliche Weise die grundsätzliche Eignung des Modells für die eingangs beschriebene Problemstellung. Ein Vergleich der gegebenen

(z_{i1} und z_{i2}) mit den geschätzten Anfrage- und Buchungshäufigkeiten (\hat{z}_{i1} und \hat{z}_{i2}) dokumentiert eine hohe Übereinstimmung, und zwar nicht nur für den Fall monoton steigender Häufigkeiten, sondern auch in den Fällen, in denen diese Monotonie durchbrochen wird, etwa bei z_{42} .

i	x_{i11}, \dots, x_{i15}	x_{i21}, \dots, x_{i25}	z_{i1}	z_{i2}	\hat{z}_{i1}	\hat{z}_{i2}
1	0 0 0 0 1	0 0 0 0 1	8	0	9,680	2,451
2	0 0 0 0 1	0 0 0 1 0	19	7	23,345	8,580
3	0 0 0 0 1	0 0 1 0 0	49	34	56,370	33,706
4	0 0 0 0 1	0 1 0 0 0	178	32	182,207	26,965
5	0 0 0 0 1	1 0 0 0 0	747	204	729,398	205,299
6	0 0 0 1 0	0 0 0 0 1	4	1	4,487	0,673
7	0 0 0 1 0	0 0 0 1 0	8	3	10,821	2,354
8	0 0 0 1 0	0 0 1 0 0	22	10	26,130	9,248
9	0 0 0 1 0	0 1 0 0 0	87	0	84,460	7,398
10	0 0 0 1 0	1 0 0 0 0	343	62	338,102	56,327
11	0 0 1 0 0	0 0 0 0 1	3	1	1,857	0,611
12	0 0 1 0 0	0 0 0 1 0	8	2	4,478	2,137
13	0 0 1 0 0	0 0 1 0 0	15	9	10,812	8,396
14	0 0 1 0 0	0 1 0 0 0	35	8	34,949	6,717
15	0 0 1 0 0	1 0 0 0 0	131	49	139,904	51,139
16	0 1 0 0 0	0 0 0 0 1	1	1	0,571	0,142
17	0 1 0 0 0	0 0 0 1 0	3	1	1,376	0,496
18	0 1 0 0 0	0 0 1 0 0	7	0	3,323	1,947
19	0 1 0 0 0	0 1 0 0 0	12	3	10,739	1,558
20	0 1 0 0 0	1 0 0 0 0	36	11	42,991	11,858
21	1 0 0 0 0	0 0 0 0 1	1	1	0,406	0,124
22	1 0 0 0 0	0 0 0 1 0	3	1	0,980	0,434
23	1 0 0 0 0	0 0 1 0 0	6	2	2,365	1,704
24	1 0 0 0 0	0 1 0 0 0	8	1	7,645	1,363
25	1 0 0 0 0	1 0 0 0 0	24	9	30,604	10,376

Tab. 3: Synthetische und geschätzte Anfrage- und Buchungshäufigkeiten

Die gute Anpassung des Modells spiegelt sich auch in der unmittelbaren Interpretierbarkeit der Parameterschätzer wieder. So schlägt sich z. B. die Proportionalität zwischen den Anfrage- und Buchungshäufigkeiten, wie Tabelle 4 zu entnehmen ist (Hinweis: Auf die in Klammern stehenden Werte wird im Anschluss eingegangen.), im Verhältnis der beiden Interzepte zueinander nieder. Darüber hinaus ist sowohl bei der Anfrage als auch bei der Buchung sehr schön die aufgrund der vorgegebenen Datenstruktur zu erwartende gegenläufige Struktur der Schätzer für die Ausprägungen des ersten und des zweiten Merkmals erkennbar. Die aufsteigenden Parameterwerte der Ausprägungen des Merkmals 1 bedeuten, dass die Wahrscheinlichkeit für eine Anfrage oder Buchung umso größer wird, je weiter rechts die 1 im Merkmalsvektor steht. Genau umgekehrt verhält es sich bei Merkmal 2. Hier wird eine Anfrage bzw. Buchung umso wahrscheinlicher je weiter links die 1 steht.

Die in Klammern angegebenen Schätzer erhält man, wenn bei der Modellkalibrierung – in exemplarischer Weise – auf die Spezifikation 14 verzichtet, diese somit quasi als (bislang) nicht existent betrachtet wird. Die sich hieraus ergebenden Unterschiede sind nur geringfügiger Natur und können somit vernachlässigt werden. Analoges gilt – bis auf Spezifikation 14 – für die korrespondierenden geschätzten Anfrage- und Beratungshäufigkeiten. Setzt man die in Klammern notierten Parameterschätzer nun zusammen mit der zu Spezifikation 14 gehörigen Merkmalskonstellation (00100, 01000) in die Modellgleichung ein, so erhält man als Prognose für die Häufigkeit einer Anfrage dieser Leistung einen Wert von $z_{14,1}^{prog} = 34,930$ und für die Buchungshäufigkeit einen Wert von $z_{14,2}^{prog} = 6,329$. Beide Prognosewerte weisen nur geringe Unterschiede zu ihren auf Basis der gesamten Datengrundlage geschätzten Pendanten aus.

Anfrage		Buchung	
$\hat{\beta}_1$	1,8325 (1,8324)	$\hat{\beta}_2$	0,9519 (0,9455)
$\hat{\beta}_{111}$	-1,1201 (-1,1200)	$\hat{\beta}_{211}$	-1,0907 (-1,0870)
$\hat{\beta}_{112}$	-0,7802 (-0,7802)	$\hat{\beta}_{212}$	-0,9571 (-0,9535)
$\hat{\beta}_{113}$	0,3997 (0,3994)	$\hat{\beta}_{213}$	0,5044 (0,4835)
$\hat{\beta}_{114}$	1,2821 (1,2822)	$\hat{\beta}_{214}$	0,6010 (0,6046)
$\hat{\beta}_{115}$	2,0510 (2,0510)	$\hat{\beta}_{215}$	1,8943 (1,8979)
$\hat{\beta}_{121}$	2,7088 (2,7088)	$\hat{\beta}_{221}$	2,4783 (2,4848)
$\hat{\beta}_{122}$	1,3217 (1,3215)	$\hat{\beta}_{222}$	0,4484 (0,4161)
$\hat{\beta}_{123}$	0,1485 (0,1485)	$\hat{\beta}_{223}$	0,6715 (0,6780)
$\hat{\beta}_{124}$	-0,7331 (-0,7330)	$\hat{\beta}_{224}$	-0,6968 (-0,6903)
$\hat{\beta}_{125}$	-1,6134 (-1,6134)	$\hat{\beta}_{225}$	-1,9495 (-1,9431)

Tab. 4: Parameterschätzer der bivariaten Poisson-Regression

Der vorliegende Ansatz bietet damit eine der Conjoint-Analyse vergleichbare Möglichkeit der Bewertung der Akzeptanz neuer Leistungen (Merkmalsausprägungskombinationen) durch den in dieser Weise modellierten Markt. Die Überprüfung, inwieweit sich dieser Ansatz auch auf reale Daten übertragen lässt, ist Gegenstand der weiteren Betrachtungen.

3.2 Analyse realer Daten aus der Tourismusbranche

Bei der folgenden Beispielanwendung konnte dankenswerterweise auf Transaktionsdaten eines bekannten Online-Reiseveranstalters zurückgegriffen werden. Der Rohdatensatz umfasst 3561 Reisebuchungen aus dem Zeitraum Herbst 2004 bis Herbst 2005. Jede Buchung wird anhand von 10 Merkmalen beschrieben. Die hierbei möglichen Ausprägungen sind der Tabelle 5 zu entnehmen.

In die Analyse gingen insgesamt 430 gemäß Tabelle 2 kodierte Spezifikationen resp. Buchungsprofile ein. Da eine eindeutige Identifikation der getätigten Anfragen auf Basis der verfügbaren Log-File-Daten nicht möglich war, konnte nur eine Analyse mittels univariater Poisson-Regression durchgeführt werden. Das entsprechende Modell erzielte gemäß Akaike Information Criterion (AIC) eine Anpassungsgüte von -1825,972. Durch eine zusätzliche A-priori-Segmentierung konnte dieser Wert noch deutlich verringert werden, was auf eine gewisse Heterogenität in den Präferenzen der Kunden schließen lässt.

Merkmale	Ausprägungen
Flugdistanz	Langstrecke, Mittelstrecke
Reisedauer	1-4 Tage, 5-13 Tage, > 13 Tage
Preis	< 55 €/pro Tag, 55-80 €/pro Tag, > 80 €/pro Tag
Unterkunft	Doppelzimmer, Apartment, Einzelzimmer
Art der Verpflegung	All-Inclusive, Halbpension, nur Frühstück, ohne Verpflegung
Hotelkategorie	bis zu 2 Sternen, 3 Sterne, 4 Sterne, 5 Sterne
Entfernung zum Strand	bis 100 m, 101-500 m, 501-1000 m, > 1000 m
Entfernung zur Stadt	bis 1 km, 1-3 km, 3-6 km, > 6 km
Hotelgröße	0-50 Zimmer, 51-100 Zimmer, 101-250 Zimmer, 251-500 Zimmer, > 500 Zimmer
Klima	subtropisch, tropisch

Tab. 5: Merkmale und mögliche Ausprägungen bei Reisebuchungen

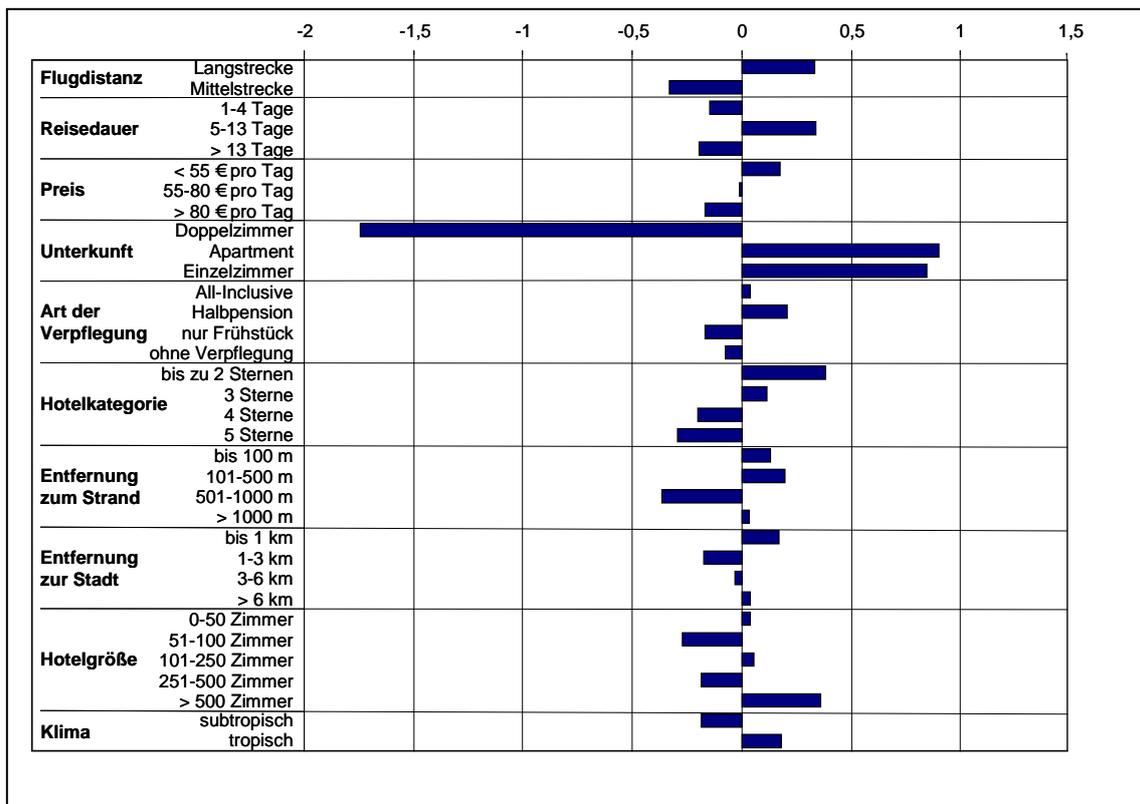


Abb. 3: Normierte Parameterschätzer auf Basis aggregierter Buchungsdaten

Aus Platzgründen sind in Abbildung 3 allerdings nur die analog zur Conjoint-Analyse als Teilnutzenwerte interpretierten Parameter der Merkmalsausprägungen für die Gesamtstich-

probe dargestellt. Dabei wurden die Parameter so normiert, dass die Summe der Teilnutzenwerte der Ausprägungen eines jeden Merkmals gleich Null ist. Dies entspricht einer üblichen Darstellung von Parametern in Conjoint-Modellen und erlaubt eine anschauliche Interpretation des Einflusses einzelner Parameter auf die Buchungshäufigkeit (Kuhfeld (2003)).

Zur Überprüfung der Signifikanz des Zusammenhangs zwischen den einzelnen erklärenden Variablen und der abhängigen Variable durchgeführte Wald-Tests zeigen, dass bei jedem Merkmal mindestens zwei Ausprägungen in einer signifikanten Beziehung ($p = 0,05$) zur Buchungshäufigkeit stehen. Im Umkehrschluss bedeutet dies, dass jedes Merkmal zur Erklärung der Buchungshäufigkeit eines Reiseangebots beiträgt.

Die vorliegenden Ergebnisse offenbaren, dass insbesondere das Merkmal „Unterkunft“ einen deutlichen Einfluss auf die Buchungshäufigkeit zu haben scheint. Auf der hier im Vordergrund stehenden aggregierten Betrachtungsebene werden Hotels mit bis zu drei Sternen bevorzugt, was zumindest teilweise mit dem Last-Minute-Charakter des betrachteten Reiseangebots zusammenhängen dürfte. Beim Merkmal „Entfernung zur Stadt“ zeigt sich, dass die Unterbringung in urbanen Gebieten (Entfernung < 1 km) den höchsten Nutzen spendet. Den zweithöchsten Nutzen generiert die Ausprägung „ > 6 km“. Die beiden anderen Ausprägungen, d. h. die Entfernungen 1-3 km und 3-6 km, liefern hingegen geringere Teilnutzenwerte. Dies lässt die Vermutung zu, dass vor allem Unterkünfte, die nahe am Stadtzentrum liegen oder Unterkünfte, die sich deutlich abseits von dicht besiedelten Gebieten („in der Natur“) befinden, bevorzugt werden. Unterkünfte, die eher in den Vorstädten oder Randgebieten (und somit in Mischzonen) liegen, werden tendenziell weniger bevorzugt. Dieses doch recht markante Präferenzmuster lässt sich u. a. damit erklären, dass unterhaltungsorientierte Urlauber eher die Nähe eines pulsierenden Stadt- und Nachtlebens suchen, während an beschaulicher Erholung interessierte Urlauber häufig eher das ruhige Hinterland bevorzugen.

Bezieht man zusätzlich das Interzept (mit einem Wert von $\hat{\beta}_0 = 1,735$) in die Betrachtungen mit ein, so lässt sich für jedes aus den vorhandenen Merkmalsausprägungen generierbare Angebotsprofil die Buchungshäufigkeit schätzen. Auf diese Weise kann auch für völlig neuartige Reisespezifikationen – ceteris paribus – eine Prognose der zu erwartenden Buchungshäufigkeit vorgenommen werden. Darüber hinaus kann mittels einfacher Simulationen überprüft werden, inwieweit Veränderungen bestehender Reiseangebote zu höheren oder niedrigeren Buchungsfrequenzen führen würden. Da häufig bestimmte Zimmerkontingente im Voraus gebucht werden (müssen), kann der Reiseveranstalter durch den Einsatz des Poisson-

Regressionsmodells die zu erwartenden Buchungshäufigkeiten besser planen und somit Negativeffekte infolge von Überbuchungen reduzieren. Aufgrund der in Abbildung 3 dargestellten Teilnutzenwerte bietet es sich für den hier betrachteten Reiseveranstalter an, dem Kunden eher ein einfaches Hotel, dafür aber die Option auf ein Einzelzimmer anzubieten. Zudem sollte die Auswahl geeigneter Partnerhotels eher durch das Merkmal „Entfernung zum Strand“ als durch das Merkmal „Entfernung zur Stadt“ beeinflusst werden. Die entsprechenden Teilnutzenwerte zeigen, dass sich eine größere Entfernung zum Strand stärker negativ auf die Buchungshäufigkeit auswirkt als eine größere Entfernung zum Stadtzentrum. Des Weiteren versprechen besonders Reisen mittlerer Dauer, d. h. ca. 1-2 Wochen, eine größere Akzeptanz seitens der Kunden. Bei einer Ausweitung des Leistungsspektrums, etwa vor dem Hintergrund der Erzielung von Kostendegressionseffekten, sollte vor allem auf Angebote dieser Dauer gesetzt werden.

3.3 Analyse realer Daten aus einem Internetportal

Ein Internetportal stellt, wie schon durch Abbildung 1 deutlich wurde, zumeist eine sehr breit gefächerte, d. h. aus mehreren Rubriken bestehende Leistung dar. Die Rubriken sind entweder inhaltlich motiviert (der so genannte Content-Bereich) oder umfassen Dienste, bei denen der Nutzer in unterschiedlichem Maß in Aktion bzw. Interaktion tritt (z. B. Routenplaner oder Online-Spiele). Das Leistungsangebot eines Portals ist i. d. R. hierarchisch aufgebaut und lässt sich typischerweise in drei Ebenen einteilen. Auf der ersten Ebene befindet sich die Startseite („Homepage“), der Ausgangspunkt des Portals, auf der die Besucher einen Überblick über das Leistungsangebot erhalten und von der aus sie sich über entsprechende Hyperlinks zu den Rubriken ihrer Wahl durchklicken können. Auf der zweiten Ebene befinden sich die zumeist thematisch gruppierten Inhaltsübersichten zu den einzelnen Rubriken, die Unter rubriken. Auf der dritten Ebene findet der Besucher dann die eigentlichen Inhalte, die ihrerseits, je nach Komplexität, wiederum unterschiedlich stark untergliedert sein können.

Für die Überprüfung unseres Modells standen uns dankenswerterweise Log-File-Daten eines führenden Internetportaltreibers zur Verfügung. Die Daten entsprechen im Wesentlichen dem Extended Common Log-File Format (ECLF) (Hippner et al. (2002, S. 10)) und repräsentieren 10922 Sessions. Bei dem betrachteten Portal gehörten zum Zeitpunkt der Datengenerierung, d. h. im Dezember 2002, Reisen, Unterhaltung, Erotik, Finanzen, Musik und Sport zu den stark frequentierten Rubriken (vgl. Tabelle 6). Diese sechs Rubriken gingen als binär kodierte Merkmale mit unterschiedlichen Anzahlen von Ausprägungen (Unterrubriken) in das

Modell ein. Die übrigen, weniger häufig frequentierten Rubriken wurden in einer synthetischen Rubrik „Sonstige“ zusammengefasst und in datenanalytischer Hinsicht wie Unterrubriken derselben behandelt. Zu den im Betrachtungszeitraum geringer frequentierten Rubriken zählen z. B. Karriere/Ausbildung, Immobilien und Partnersuche.

Rubrik	Unterrubrik	Anzahl Besuche	Rubrik	Unterrubrik	Anzahl Besuche
Home	Homepage	6694	Finanzen	Finanzen (allg.)	340
Reisen	Reisen	284		Finanzen (sonst.)	250
	Routenplaner	400	Musik	Musik (Index)	372
	Übersetzungsprog.	563		Künstler	116
	Flugangebote	464		Karaoke	31
	Sonderangebote	46		Besonderheiten	41
	Reisen (sonst.)	221		MP3	124
Unterhaltung	Unterhaltung (Index)	534	Musik (sonst.)	194	
	Fernsehen	30	Sport	Sport (Index)	105
	Witze	122		Fußball	56
	Horoskop	428		Service	10
	Zeitvertreib	1261		Sport (allgemein)	143
	Unterhaltung (sonst.)	1574		Sport (sonst.)	20
Erotik	Erotik (Index)	2102		Sonstige	Auto/Motor
	Bilder-Galerie	1779	Karriere/Ausbildung		194
	Video	377	Computer/Internet		403
	Fun	188	Informationen		280
	Allgemeines	2778	Lifestyle/Gesundheit		283
	Erotik (sonst.)	178	Nachrichten		49
Finanzen	Finanzen (Index)	277	Partnersuche		56
	Aktien	119	Immobilien		73
	Pers. Finanzen	143	Wetter		162
	Finanznachrichten	32	Sonst. Angebote		1586

Tab. 6: In die Analyse einbezogene Rubriken und Unterrubriken des Internetportals

Um die am Anfang und am Ende des betrachteten Zeitfensters identifizierten Sessions nicht zeitlich zu zensieren (Blossfeld und Rohwer (1995)), wurde vor und nach diesem Zeitintervall jeweils eine Pufferzone von 30 Minuten eingerichtet. Sessions, die ganz oder teilweise in diese Pufferzonen fielen, wurden von den Analysen ausgeschlossen. Auch jene Sessions, bei denen nur die Startseite des Portals besucht wurde, blieben aufgrund ihrer geringen Aussagekraft für den vorliegenden Analysefokus außer Acht. Nach Bereinigung der Daten standen insgesamt 10912 vollständige Sessions für die weitere Analyse zur Verfügung.

In Abbildung 4 sind ausgewählte, normierte Parameterschätzer der einzelnen Merkmalsausprägungen dargestellt. Werte größer Null zeigen einen im Mittel positiven Einfluss der betreffenden Merkmalsausprägung auf die Verweildauer an, während negative Werte auf einen gegenteiligen Effekt hindeuten. Man erkennt, dass insbesondere Unterrubriken aus den Berei-

chen Erotik, Finanzen und Musik zu längeren Verweildauern Anlass geben, den Besuchern also offensichtlich einen vergleichsweise hohen Nutzen spenden.

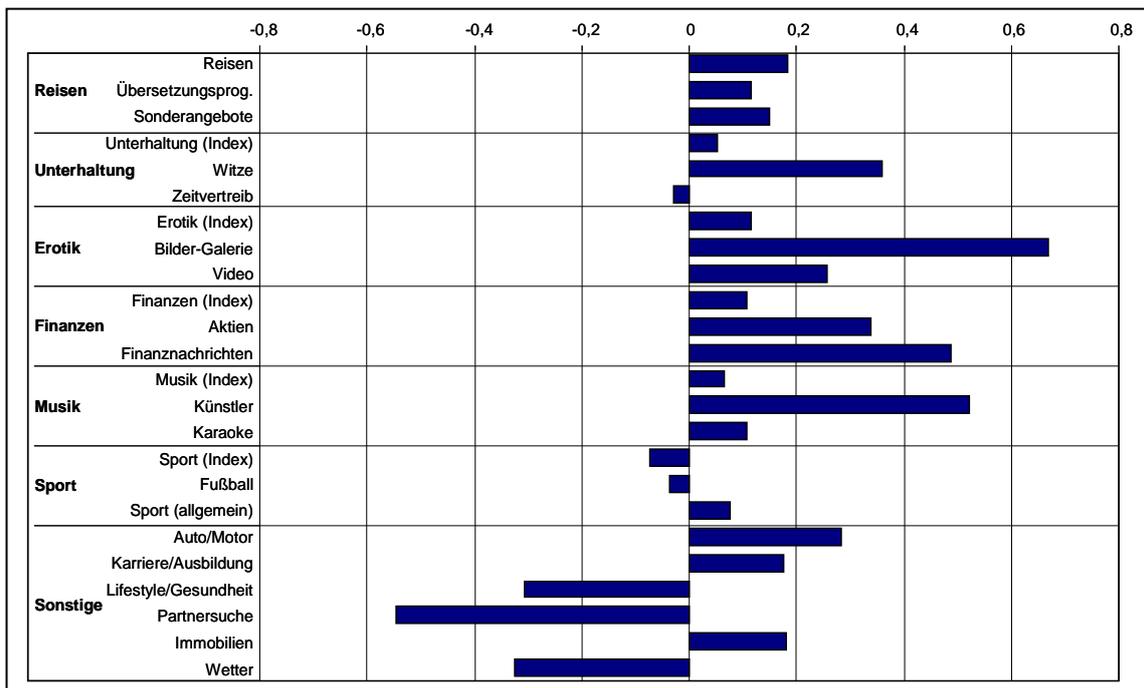


Abb. 4: Normierte Parameterschätzer für das Portalbeispiel

Stellt man die in Abbildung 4 dargestellten Parameterschätzer den in Tabelle 6 angegebenen Besuchsfrequenzen je Rubrik bzw. Unterrubrik gegenüber, so wird der Mehrwert der modellbasierten Analyse deutlich. Die Rubrik „Immobilien“ weist z. B. nur eine vergleichsweise geringe Besucherfrequenz auf und zählt somit zu den weniger häufig genutzten Inhalten des Portals. Gleichzeitig kommt aber durch den zugehörigen Parameterschätzer 0,18 der grundsätzlich positive Einfluss dieser Rubrik auf die Verweildauer im Portal zum Ausdruck. Besucher, die Seiten der Rubrik „Immobilien“ aufgerufen haben, nutzten das Portal im Mittel länger als Personen, die diese Seiten nicht besuchten. Die Inhalte dieser Rubrik spendeten den Besuchern offensichtlich einen merklichen Nutzen und verdienten es deshalb, auch weiterhin gepflegt zu werden. Die Rubrik „Partnersuche“ verfügt ebenfalls nur über eine geringe Besuchsfrequenz und zeichnet sich darüber hinaus durch einen negativen Parameterwert aus. Die Inhalte der betreffenden Seiten scheinen also zum Zeitpunkt der Datenerhebung eher zu einer Verkürzung der Besuchszeiten geführt zu haben. Damit stellt sich natürlich die Frage nach den Ursachen für diesen Effekt. Ein möglicher Grund könnte der Aufruf verlinkter externer Internetangebote sein, die nicht zum Portal gehörten und somit datentechnisch nicht erfasst wurden. Der identifizierte Effekt kann in diesem Fall sogar positiver Natur gewesen sein, nämlich z. B. dann, wenn es sich bei diesen Links um Bannerwerbung handelte und diese dem

Portalbetreiber je Klick vergütet wurden. Ein anderer Grund könnte sein, dass Besucher der Rubrik „Partnersuche“ nicht die Inhalte vorfanden, die sie dort erwarteten. Möglicherweise vermuteten einige der Besucher hier auch erotische Angebote, was auf einen Verbesserungsbedarf hinsichtlich der Benutzerführung bzw. der Kommunikation der Inhalte hindeutet. Maßnahmen in dieser Richtung müssen eine eindeutige Identifizierbarkeit der betreffenden Inhalte sicherstellen und einen schnellen Zugang ermöglichen. Die bereitgestellten Angebote können aber natürlich auch bewusst aufgerufen worden sein, entsprachen dann jedoch nicht den Ansprüchen der Besucher. In diesem Fall wäre eine Überarbeitung der angebotenen Inhalte angeraten.

Bei anderen Rubriken und Unterrubriken stellt sich die Sache genau andersherum dar. Die Unterrubrik „Zeitvertreib“ beispielsweise weist eine vergleichsweise hohe Besuchszahl auf. Der Wert des entsprechenden Parameters fällt hingegen eher klein aus. Dies bedeutet, dass Besucher, die diese Unterrubrik aufsuchten, das Portal im Mittel schon nach kurzer Zeit wieder verließen. Dieses Phänomen ist vor allem deshalb bemerkenswert, als davon ausgegangen werden kann, dass für das Gros der Besucher, die dieses Angebot aufsuchten, die schnelle Suche und Aufnahme von Informationen wohl eher nicht im Vordergrund stand.

Bereits diese wenigen Beispiele lassen erkennen, dass die modellbasierte Herangehensweise einen echten Mehrwert gegenüber der reinen Betrachtung von Besuchshäufigkeiten bieten kann. Der Einsatz des vorgeschlagenen Poisson-Regressionsmodells ermöglicht eine differenzierte Betrachtung, die explizit auf die Auswirkungen der vorgefundenen Inhalte in Form von Attraktions- bzw. Bindungspotenzialen abzielt. Für eine umfassende Bewertung der angesprochenen Phänomene sind allerdings weiterführende Analysen erforderlich, die auch qualitative Aspekte mit einbeziehen. Der vorgeschlagene Ansatz kann aber dabei helfen, herauszufinden, welche Bereiche die Aufenthaltsdauer im Portal zukünftig erhöhen können und wo sich möglicherweise Verbesserungsbedarf abzeichnet. So sollte sich der Portalbetreiber im vorliegenden Fall z. B. Gedanken über die vergleichsweise geringen Nutzenwerte beim Sportangebot machen. Für eine abschließende Analyse müsste allerdings auch das verfügbare Angebot der Wettbewerber Berücksichtigung finden. In Deutschland gibt es zahlreiche kostenfreie Online-Sportinformationsdienste, die ein umfassendes und hoch aktuelles Leistungsangebot präsentieren. Daneben legen aber auch die Unterrubriken „Lifestyle/Gesundheit“ und „Partnersuche“ eine Überarbeitung nahe. Möglicherweise sollte sogar in Erwägung gezogen werden, diese Unterrubriken ganz zu schließen, und stattdessen eine stärkere Fokussierung des Angebots anzustreben. Eine solche „Angebotsverschlinkung“ könnte insbesondere dann

positive Auswirkungen auf die Gesamtbeurteilung des Portals haben, wenn davon auszugehen ist, dass die Besucher dieser Unterrubriken tatsächlich einen negativen Nutzen empfinden (etwa im Sinne von Zeitverschwendung), z. B. durch eine entstandene Desorientierung oder unzureichende Inhalte. Eine Fokussierung könnte auch im Hinblick auf die Vermarktung von Werbeplätzen von Vorteil sein, da die durch die betreffenden Inhalte attrahierte Nutzerschaft sehr wahrscheinlich homogener ausfallen und somit die Ansprache der Werbetreibenden geringere Streuverluste aufweisen würde.

Auch die Resultate zur Unterrubrik „Wetter“ verdienen eine genauere Betrachtung. Das Lesen eines Wetterberichts stellt im Gegensatz zur Inanspruchnahme stärker ausdifferenzierter Leistungsangebote i. d. R. eine Nutzung dar, die zeitlich stark begrenzt ist. Nachdem der Besucher die ihn interessierenden Wetterdaten abgerufen hat, ist sein Informationsbedarf zumeist befriedigt. Für einen längeren Aufenthalt in dieser Unterrubrik besteht somit kaum Veranlassung. Der wahrgenommene Nutzen des Portals als Ganzes kann somit erst durch das Aufsuchen anderer (Unter-) Rubriken weiter erhöht werden. Folglich gilt es zunächst zu eruieren, ob sich die Besucher der Unterrubrik „Wetter“ tatsächlich zum überwiegenden Teil auf das Abrufen von Wetterinformationen beschränken oder sie eben auch weitere Leistungsangebote des Portals in Anspruch nehmen. Im ersten Fall erweist es sich als vorteilhaft, dass die wiederholte, aus eigenem Antrieb heraus erfolgende Auseinandersetzung mit einem Angebot eine positivere Wahrnehmung desselben hervorrufen kann. Dieses Phänomen wird in der Literatur unter dem Begriff des „Mere Exposure-Effekt“ diskutiert (siehe hierzu auch Decker et al. (2005)). Darüber hinaus können die aus einer Modellanwendung resultierenden Erkenntnisse hinsichtlich des Bindungspotenzials einzelner (Unter-) Rubriken z. B. bei der Vermarktung entsprechender Werbeplätze von Nutzen sein.

4. Fazit und Ausblick

Im vorliegenden Beitrag wurde ein einfaches Modell zur entscheidungsunterstützenden Analyse von Nachfragedaten aus Internetplattformen auf Basis der Poisson-Regression vorgestellt. Dieses ermöglicht, unter Rückgriff auf die infolge der Inanspruchnahme von Internetangeboten zwangsläufig anfallenden Log-File-Daten, den Erhalt aussagekräftiger Informationen über die Attraktivität und Adäquanz bestehender Leistungsangebote. Der vorgeschlagene Ansatz lässt Parallelen zur weit verbreiteten Conjoint-Analyse erkennen und ermöglicht für die Merkmale eines Leistungsangebotes (z. B. die verschiedenen Elemente einer Pauschalreise oder die einzelnen Rubriken eines Internetportals) die Schätzung von Parametern, die sich

im weiteren Sinne als Teilnutzenwerte interpretieren lassen. Ihre inhaltliche Auswertung kann u. a. bei der kundenorientierten Evaluation geplanter Neuprodukte (etwa in Form neuer Reiseprofile oder neuer Rubrikinhalte) von Nutzen sein, ohne aber im Gegenzug zusätzliche Erhebungskosten wie bei einer Conjoint-Analyse hervorzurufen.

Für Unternehmen bietet der Einsatz des hier vorgestellten Modells eine kosteneffiziente Möglichkeit, um zusätzliche Informationen über die Präferenzstrukturen ihrer Kunden zu erhalten. Das Modell trägt damit zu einem vertieften Verständnis des Konsumentenverhaltens im Internet bei. Online-Anbieter von Produkten und Dienstleistungen können hierdurch ihr Produktangebot optimieren sowie neue Angebote entwickeln. Da nach wie vor viele Kunden Produktkonfiguratoren aufgrund der damit verbundenen Komplexität negativ bewerten bzw. erst gar nicht verwenden (Dellaert und Stremersch (2005)), stellt das Modell eine Alternative zur Identifizierung Erfolg versprechender Standardangebote dar.

Neben diesen offenkundigen Vorteilen seien an dieser Stelle aber auch die Grenzen dieses Ansatzes aufgezeigt: So können damit natürlich nur für jene Leistungsausprägungen Präferenzen bestimmt werden, die tatsächlich am Markt verfügbar sind. Des Weiteren basiert der Ansatz auf Daten eines einzelnen Online-Anbieters. Die auf diese Weise zu erhaltenen Präferenzen können sich folglich von denen der Nutzerschaft anderer Anbieter unterscheiden.

Abschließend bleibt zu hoffen, dass der vorliegende Beitrag Anstöße für die Entwicklung weiterer, modellbasierter Ansätze zur Erforschung von Konsumentenpräferenzen auf Basis von Log-File-Daten gibt und damit das Werkzeugspektrum des Web Usage Mining in zweckdienlicher Weise ergänzt (Teltzrow (2005)). Auch die Einbindung des vorgestellten Ansatzes in touristische Marketinginformationssysteme ist denkbar und sollte mit vertretbarem Aufwand realisierbar sein (Wöber (2002)).

Eine vielversprechende Erweiterung des vorgestellten Ansatzes stellt die explizite Berücksichtigung der Unterschiedlichkeit des individuellen Nutzungsverhaltens mittels einer geeigneten Verteilung dar. Entsprechende Vorgehensweisen haben in der stochastischen Kaufverhaltensmodellierung bereits eine lange Tradition (vgl. hierzu etwa Decker und Wagner (2002)). Durch eine solche Modellerweiterung kann auf elegante Weise dem in Unterabschnitt 3.2 angesprochenen Problem der Heterogenität individueller Präferenzen Rechnung getragen werden.

5. Literatur

- Backhaus, K.; T. Hillig; R. Wilken (2007): Predicting Purchase Decisions with Different Conjoint Analysis Methods: A Monte Carlo Simulation, *International Journal of Market Research*, Vol. 49, No. 3, 341-364.
- Bauer, H. H.; M. M. Neumann; J. Rösger (2004): *Konsumentenverhalten im Internet*, München, Vahlen.
- Blossfeld, H.-P.; G. Rohwer (1995): *Techniques of Event History Modeling*, Mahwah, Lawrence Erlbaum Associates.
- Burklen, S.; P. J. Marron; S. Fritsch; K. Rothermel (2005): User Centric Walk: An Integrated Approach for Modeling the Browsing Behavior of Users on the Web, in: *Proceedings of the 38th Annual Symposium on Simulation*, IEEE Computer Society, 149-159.
- Cheung, C. M. K.; L. Zhu; T. Kwong; G. W. W. Chan; M. Limayem (2003): Online Consumer Behavior: A Review and Agenda for Future Research, in: *Proceedings of the 16th Bled eCommerce Conference*, Bled, 194-218.
- Constantinides, E. (2004): Influencing the Online Consumer's Behavior: The Web Experience, *Internet Research*, Vol. 14, No. 2, 111-126.
- Cooley, R.; B. Mobasher; J. Srivastava (1999): Data Preparation for Mining World Wide Web Browsing Patterns, *Journal of Knowledge and Information Systems*, Vol. 1, No.1, 5-32.
- Dahan, E.; V. S. Srinivasan (2000): The Predictive Power of Internet-Based Product Concept Testing Using Visual Depiction and Animation, *Journal of Product Innovation Management*, Vol. 17, No. 3, 99-109.
- Decker, R.; R. Wagner (2002): *Marketingforschung – Methoden und Modelle zur Bestimmung des Käuferverhaltens*, München, Moderne Industrie.
- Decker, R.; S. W. Scholz; R. Wagner (2005): Mining Consumption and Nutrition Patterns for Marketing Communication, in: Zani, S.; A. Cerioli (Eds.): *Classification and Data Analysis 2005*, Parma, Monte Università Parma, 91-94.
- Dellaert, B. G. C.; S. Stremersch (2005): Marketing Mass-Customized Products: Striking a Balance between Utility and Complexity, *Journal of Marketing Research*, Vol. 42, No. 2, 219-227.

- Drèze, X.; F. Zufryden (1997): Testing Web Site Design and Promotional Content, *Journal of Advertising*, March/April, 77-91.
- Drèze, X.; F. Zufryden (1998): A Web-based Methodology for Product Design Evaluation and Optimisation, *Journal of the Operational Research Society*, Vol. 49, September, 1034-1043.
- FUR (2007): Die 37. Reiseanalyse RA 2007, <http://www.fur.de>.
- Hippner, H.; M. Merzenich; K. D. Wilde (2002): *Handbuch Web Mining im Marketing*, Braunschweig, Vieweg.
- Kuhfeld, W. F. (2003): *Marketing Research Methods in SAS: Experimental Design, Choice, Conjoint, and Graphical Techniques*, Cary NC, SAS Institute Inc.
- Kuom, M.; B. Oertel (1999): Virtual Travel Agencies, *Netnomics*, Vol. 1, 225-235.
- Maddala, G. S. (1983): *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press.
- Meißner, M.; S. W. Scholz; R. Decker (2007): Analytic Hierarchy Process vs. Adaptive Conjoint Analysis – An Empirical Comparison, in: Preisach, C.; H. Burkhardt; L. Schmidt-Thieme; R. Decker (Eds.): *Data Analysis, Machine Learning, and Applications*, Heidelberg, Springer, forthcoming.
- Meißner, M.; S. W. Scholz; R. Wagner (2008): *Marketing Research Using Multimedia Technologies*, to appear in: Pagani, M. (Ed.): *Encyclopedia of Multimedia Technology and Networking*, 2nd Ed., Idea Group Publishing.
- Montgomery, A. L.; S. Li; K. Srinivasan; J. C. Liechty (2004): Modeling Online Browsing and Path Analysis Using Clickstream Data, *Marketing Science*, Vol. 23, No. 4, 579-595.
- Oertel, B.; T. Feil; S. L. Thio (2003): *Neue IuK-Technologien und ihre Relevanz für die Wettbewerbsfähigkeit touristischer Destinationen. Trends, Daten und Fakten im internationalen Vergleich*, Werkstattbericht Nr. 56 – März 2003, Berlin. IZT – Institut für Zukunftsstudien und Technologiebewertung.
- Rudolph, T.; S. Busch; T. Wagner (2002): Kaufbarrieren im Online-Handel aus Kundensicht, in: Ahlert, D.; R. Olbrich; H. Schröder (Hrsg.): *Jahrbuch Handelsmanagement 2002 – Electronic Retailing*, Frankfurt, Deutscher Fachverlag, 143-160.

- Skiera, B.; S. Gensler (2002): Berechnung von Nutzenfunktionen und Marktsimulationen mit Hilfe der Conjoint-Analyse, Teil 1, Wirtschaftswissenschaftliches Studium, Vol. 31, 200-206.
- Spiliopoulou, M. (2000): Web Usage Mining for Web Site Evaluation, Communications of the ACM, Vol. 43, No. 8, 127-134.
- Steckel, J. H.; R. S. Winer; R. E. Bucklin; B. G. C. Dellaert; X. Drèze, G. Häubl, S. D. Jap; J. D. Little; T. Meyvis; A. L. Montgomery; A. Rangaswamy (2005): Choice in Interactive Environments, Marketing Letters, Vol. 16, No. 3-4; 309-320.
- Tao, Y.-H.; T.-P. Hong; Y.-M. Su (2006): Improving Browsing Time Estimation with Intentional Browsing Data, International Journal of Computer Science and Network Security, Vol. 6, No. 12, 35-39.
- Teltzrow, M. (2005): A Quantitative Analysis of E-Commerce: Channel Conflicts, Data Mining, and Consumer Privacy, Dissertation, Humboldt-Universität zu Berlin.
- Tutz, G. (2000): Die Analyse kategorialer Daten – Anwendungsorientierte Einführung in Logit-Modellierung und kategoriale Regression, München, Oldenbourg.
- Wöber, K. (2002): Das Internet als Transportmittel touristischer Marktforschungsinformationen am Beispiel von TourMIS, Tourismus Journal , Vol. 6, No. 1, 25-48.