Prof. Dr. Silke Meyer und
Dipl.-Math. oec. Alexander Börsch
arbeiten an der HS Mittweida mit Forschungsschwerpunkten Business Intelligence/Data Mining.
E-Mail: meyer@hs-mittweida.de

Dr. Michael Thess ist Vorstand der prudsys AG, Chemnitz.
E-Mail: thess@prudsys.de

Lernen in Echtzeit

Empfehlungssysteme zur Preisoptimierung

Data-Mining-Verfahren zur Nutzbarmachung großer Datenmengen finden seit Jahren zur Analyse des Kundenverhaltens Anwendung, um beispielsweise die Kundengewinnung, Betrugserkennung oder Kündigerprävention zu unterstützen. Für das Cross- und Up-Selling werden in allen großen Online-Shops Empfehlungssysteme (englisch: recommender systems) eingesetzt. Der Beitrag stellt selbstlernende Empfehlungssysteme vor und diskutiert deren algorithmischen Erweiterungen (Gegenstand des vom BMBF geförderten Verbundprojekts „PriMini", FKZ:1717X08) für die aktive, dynamische Preissteuerung im Handel.

Preisoptimierung im Handel

Ein brisantes Thema, nicht nur für Handelsunternehmen, ist die optimale Gestaltung von Verkaufspreisen. Auf Grund der wachsenden Komplexität externer und interner Einflussfaktoren sowie steigender Transaktionsgeschwindigkeiten ist diese Aufgabe kaum noch manuell durchführbar. Beschleunigt durch die derzeitige Wirtschaftssituation wird die algorithmische Unterstützung des Yield Managements, der dynamischen Preis- und Kapazitätssteuerung, in der Praxis stark nachgefragt und mit hohen Potenzialen bemessen, allerdings bislang im Online-Handel kaum umgesetzt [BeS08].
Technologien zur Echtzeitverarbeitung umfangreicher Datenmengen ermöglichen bereits heute gezielte Produkt- oder auch Preisempfehlungen in Online-Shops. Auch im stationären Handel ist dies basierend auf einem hohen Datenintegrationsstand mittels POI-Systemen (zum Beispiel Multimedia-Waagen, Terminals) oder elektronischer Preisschilder realisierbar. Die Schwierigkeit liegt darin, zu einem Produkt entsprechend seiner Lebenszyklusphase den „optimalen" Preis zu finden. Ein Optimum ist unter meist konträren Zielstellungen des Anwendungsszenarios (zum Beispiel Umsatzmaximierung versus Minimierung der Lagerbestände) nicht eindeutig bestimmbar. Daher besteht die Herausforderung darin, den Preis als Pareto-Optimum in einem vorgegebenen Preisintervall (abgeleitet von Kosten und Konkurrenzpreisen) adäquat zu wählen.

Reinforcement Learning in Empfehlungssystemen

Für Produktempfehlungen wurden bisher häufig die Methoden des sogenannten Collaborative Filtering und Content-based Filtering kombiniert. Während die erste Methode lediglich die Transaktionsdaten („Warenkörbe") der Vergangenheit analysiert, nutzt das zweite Verfahren Ähnlichkeiten der Artikel und unterstützt die gezielte Empfehlung von Neuprodukten.
Einen besonders interessanten Ansatz für Empfehlungssysteme liefert das „Reinforcement Learning", das erstmals kommerziell von der prudsys AG, Chemnitz, in einer Empfehlungsmaschine, der prudsys RE (Recommendation Engine), umgesetzt wurde. Die Basis der prudsys RE bildet die Xelopes-Bibliothek von höchst performanten Data-Mining-Algorithmen (vgl. auch [The01]).
Die Methode des Reinforcement Learning (RL) basiert auf der Dynamischen Programmierung. Ein „selbstlernender Agent" lernt dabei anhand von Belohnung R und Bestrafung (negative Belohnung) eine Nutzenfunktion (vgl. [Sut98]). Er verfolgt dabei nicht die Strategie im jeweils nächsten Schritt ein „Optimum" zu erreichen, sondern über mehrere Schritte die insgesamt nutzenmaximierende Lösung zu realisieren. Hierzu nutzt der RL-Agent Entscheidungsregeln, die jedem Zustand des definierten Zustandsraumes eine empfohlene Aktion a_t zuordnen, zum Beispiel eine Produktempfehlung zu einem durch den Kunden recherchierten Artikel.
Während der fortlaufenden Nutzung des Systems lernt der Agent aus den Kundenreaktionen (das heißt aus der Annahme / Ablehnung von Empfehlungen), welche Zustand-Aktions-Tupel (s,a) besonders nutzbringend sind. Hierzu

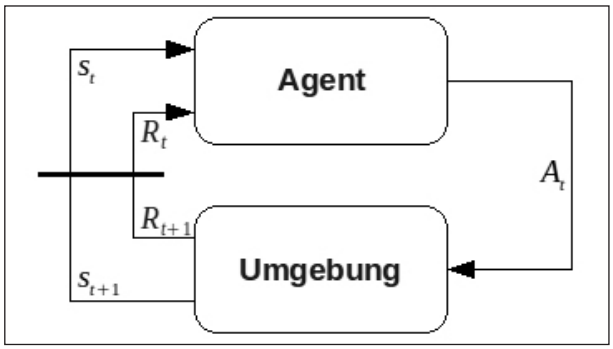

Abb. 1: *Reinforcement Learning basierend auf der Dynamischen Programmierung. Es schlägt – abhängig von Regeln – Aktionen vor.*

wird die Aufgabe des Reinforcement Learning zu einem Markov-Prozess abstrahiert, wobei die Markov-Eigenschaft: „Die zukünftigen Ereignisse hängen nur vom aktuellen Zustand s ab“ in der Praxis nicht absolut, aber hinreichend erfüllt ist. Unter dieser Annahme, wird die zu lernende Nutzenfunktion V^{π} durch die Bellmann-Gleichung beschrieben, welche hier exemplarisch für diskrete Probleme angegeben wird (vgl. auch [Bel52], [Sut98]):

$$V^{\pi}(s) = \sum_{a} \pi(s,a) \sum_{s'} P^{a}_{ss'} \left\lfloor R^{a}_{ss'} + \gamma V^{\pi}(s') \right\rfloor$$

Dabei repräsentiert s den aktuellen Zustand und a die Aktion, um den nächsten Zustand s´ mit einer Übergangswahrscheinlichkeit $P^{a}_{ss'}$ zu erreichen. $R^{a}_{ss'}$ kennzeichnet die zugehörige Belohnung. Die Bellmann-Gleichung, welche den Nutzen V^{π} als kumulierten Wert der Belohnungen über die Zeit (mit Berücksichtigung eines Verminderungsfaktors γ) beschreibt, kann unter anderem mittels dem Verfahren der Wert-Iteration approximativ gelöst und damit die sogenannte Policy π (s,a) gelernt werden. Besonderheiten, welche bei der Übertragung des RL-Ansatzes auf die Aufgabe der Preisoptimierung auftreten, werden nachfolgend diskutiert.

Dynamische Ansätze zur Preisoptimierung

Für eine Gewinn- beziehungsweise Deckungsbeitragsoptimierung nimmt die produktspezifische Preis-Absatz-Funktion eine zentrale Rolle ein. Der Absatz hängt selbst zwar vom Preis, jedoch nicht alleinig von diesem Einflussfaktor ab. Daher darf die Approximation der Preis-Absatz-Funktion nicht als über die Zeit konstant angesehen werden. Aus diesem Grund werden dynamische Algorithmen benötigt, welche ihr zugrunde liegendes Modell adaptiv in Echtzeit anpassen, um sich schnell auf Veränderungen des Kaufverhaltens der Kunden und so indirekt auch der Konkurrenzpreise einzustellen.
Dies ist mittels des Reinforcement Learning möglich, indem der erzielte Gewinn- oder Deckungsbeitrag als Belohnung $R^{a}_{ss'}$ Verwendung findet [Bör07]. Eine Aktion a des RL-Agenten ist dann die Preisänderung, wobei der Zustand s dem aktuellen Preis entspricht. Die Abbildung 2, angelehnt an [SuB98], verdeutlicht die Grundidee der dynamischen Preisoptimierung, wobei die Preise in einem vorgegebenen Intervall variiert werden.
Die Anwendung des RL-Verfahrens erfordert, dass alle Zustände s des Zustandsraumes besucht werden, um eine Aussage über sie treffen zu können. Das würde bedeuten, dass jeder Preis einmal realisiert werden müsste. Eine Möglichkeit dies zu umgehen, eröffnet sich durch den Ansatz, die Nutzenfunktion nicht vollständig diskret, sondern approximativ zu bestimmen [TsR97]. Dies kann mittels einer Linearkombination geeigneter Ansatzfunktionen aus dem Bereich der Finiten-Elemente-Methode

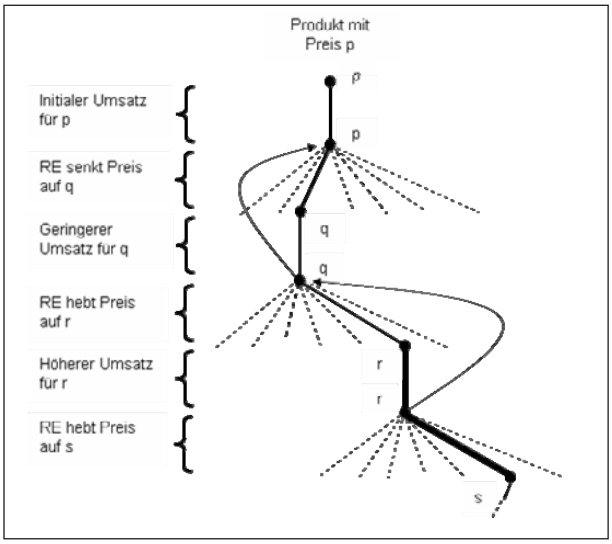

Abb. 2: Grundidee der dynamischen Preisoptimierung

realisiert werden, wobei die Herausforderung darin besteht, dies nicht zu grob und dennoch hinreichend detailliert vorzunehmen, um mit einer geringen Streuung an Werten gute Ergebnisse zu erzielen. Des Weiteren sollte diese Methode initialisiert werden, da andernfalls viele Iterationen nötig sind, um ein hinreichend gutes Approximationsmodell zu errechnen.

Datenvorverarbeitung unter Berücksichtigung kategorischer Daten

Die Lösungsqualität eines jeden Data-Mining-Modells hängt entscheidend von der Datenvorverarbeitung ab, die auch für die Preisoptimierung mit größter Sorgfalt durchzuführen ist. Dabei kommt es besonders darauf an, alle potenziellen Einflussgrößen, auch nicht-numerische Merkmale adäquat zu berücksichtigen. Um kategorische Daten in geeigneter Weise in einen Approximationsalgorithmus einfließen zu lassen, gibt es verschiedene Möglichkeiten. Eine Methode, welche in der Praxis meist verwendet wird, ist das Binarisieren der Daten. Allerdings ist dieses Vorgehen bei Merkmalen mit zahlreichen Ausprägungen nicht mehr sinnvoll, da zum einen die Daten „aufgebläht“ werden und zum anderen wertvolle Informationen über Ähnlichkeiten der Merkmalsausprägungen verloren gehen.
Um dies zu vermeiden, gibt es eine andere in der Praxis verwendete Klasse von Methoden, die jeder Ausprägung eines Merkmals einen Zahlenwert zuordnet. Die einfachste Methode dieser Art ist das „Häufigkeitsbasierende Ersetzen“. Es gibt auch komplexere Verfahren, bei denen der Zahlenwert einer Ausprägung in Abhängigkeit vom Zielmerkmal bestimmt wird. Hierbei gehen allerdings ebenfalls Informationen verloren, etwa die Beziehung der Ausprägungen untereinander. Um diesen Informationsverlust zu minimieren, gibt es eine dritte Klasse von Methoden, welche lediglich voraussetzen, dass eine

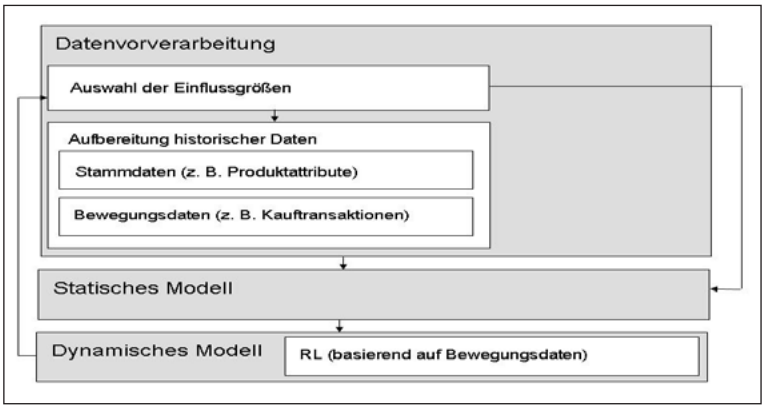

Abb. 3: Aus einem statischen Modell wird ein dynamisches

Metrik auf dem Merkmalsraum definiert ist - ein in der Praxis nicht zu unterschätzender Vorteil.
Eine Metrik lässt sich auf den meisten Produktattributen sinnvoll definieren, beispielsweise auf der Menge der Zeichenketten durch den Levenstein-Abstand. Derartige Metriken sind wissensintensiv, das heißt es sollte unternehmensspezifisches Fachwissen in deren Definition einfließen. Gelingt dies, so gibt es Approximationsverfahren, welche nach geringen Modifikationen mit diesen Metriken arbeiten und damit auch nicht-numerische Merkmale berücksichtigen können. Besonders geeignet sind das Kriging-Verfahren und die sogenannte Shepard-Methode [HoL89].

Umsetzung der dynamischen Preisoptimierung

Mit einem in dieser Weise erzeugten statischen Modell lässt sich nun ein dynamisches Modell initialisieren. Für die Anwendung des Reinforcement Learning geschieht dies, indem die Umgebung, in welcher der Agent lernt, zunächst durch das statische Modell ausgetauscht wird. Somit kann der Agent die Informationen des statischen Modells aufnehmen und das RL-Modell entsprechend initialisieren. Danach wird die Umgebung durch die Realität ersetzt (zum Beispiel durch entsprechende Preisänderungen und das jeweilige Kunden-[Klick-/Kauf-]verhalten) und der Agent kann sein Modell adaptiv anpassen.

Das Reinforcement Learning wird seit langem erfolgreich von der prudsys AG, Chemnitz, für Produktempfehlungen eingesetzt. Im Allgemeinen werden dabei in großen Shops mehrere Millionen Entscheidungsregeln bei einigen zehntausend Produkten gelernt und eine Umsatzsteigerung von 5 - 15 % erreicht. Aufbauend auf diesen Erfolgen werden die Algorithmen derzeit erweitert, so dass sie zusätzlich zu den Empfehlungen von Produkten zu Produkten, ebenfalls wie beschrieben dynamische Preisvariationen vorschlagen.
Diese Zielstellung der Gewinn- beziehungsweise Deckungsbeitragsoptimierenden Preisoptimierung wird derzeitig für den Abverkauf von Produkten nachgefragt. Erfahrungen werden seit Juni 2009 in einem Online-Shop gesammelt. Dabei zeigen die ersten Ergebnisse, dass es möglich ist, algorithmisch gestützt den Preis produktspezifisch „optimal“ abzusenken. Preissenkungen erfolgen damit nicht mehr nach der „Rasenmähermethode“, das heißt prozentual einheitlich über das gesamte Abverkaufssortiment. Die dynamische Preisoptimierung wird hierbei in erster Linie durch die Lernfähigkeit der RL-Algorithmen erreicht.

Literatur:

[Bel52] Bellman, R.: On the Theory of Dynamic Programming, Proceedings of the National Academy of Sciences, 1952.
[BeS08] Berswordt, F. von; Seiler, M. O.: Chancen statt Konflikte. In: e-Commerce Magazin, 03, 2008, S. 40 – 42.
[Bör07] Börsch, A.: Dynamische Preisoptimierung im Handel, Diplomarbeit, TU Chemnitz, 2007
[SuB98] Sutton, S.; Barto, A.: Reinforcement Learning An Introduction, The MIT Press, 1998
[HoL89] Hoschek, J.; Lasser, D.: Grundlagen der geometrischen Datenverarbeitung, B.G. Teubner, 1989
[TsR97] Tsitsiklis, J. N.; Roy, B. V.: An Analysis of Temporal-Difference Learning with Function Approximation. In: IEEE Transactions on Automatic Control 42, 1997, p. 674 – 690.