

# Personalisierung internetbasierter Handelsszenarien

Michael Onken  
Matthias Pretzer  
Guido Zendel

Carl von Ossietzky Universität Oldenburg  
Fachbereich Informatik  
Abteilung Informationssysteme  
Prof. Dr. Appelrath

## 1 Motivation

In der heutigen Zeit werden immer mehr Daten von Unternehmen und anderen Organisationen gesammelt [Heg99]. Diese Daten bergen ein enormes „Wissenspotential“: Händler wollen dies z.B. nutzen, um ihre Kunden besser verstehen und an sich binden zu können. Dazu müssen sie Daten über das Kaufverhalten der Kunden sammeln und diese analysieren. Sowohl Händler als auch Kunden können von dieser Analyse in Form von personalisierten Angeboten profitieren.

Vor diesem Hintergrund wurde an der Universität Oldenburg zum WS 02/03 die Projektgruppe (zweimestrige Lehrveranstaltung mit bis zu zwölf Studierenden) *Personalisierung internetbasierter Handelsszenarien* initiiert. Ihr Ziel ist es, ein System zu konzipieren und zu realisieren, das es ermöglicht, internetbasierte Handelsplattformen mit einem auf den jeweiligen Kunden zugeschnittenen Angebot auf der Basis verteilter und temporaler Datenquellen auszustatten.

## 2 Grundlagen

Ein Großteil der anfallenden Datenmengen wird in Datenbanksystemen gespeichert. Neben den operativen Systemen, mit denen das Tagesgeschäft abgewickelt wird, haben sich Technologien etabliert, die gezielt die Analyse der Daten unterstützen [HK01]. Oft dienen hierbei Data Warehouses [BE01] als Speicherkomponente, auf deren Datenbestand Techniken des Knowledge Discovery in Databases (KDD) angewendet werden, um Wissen aus den Daten zu gewinnen [Küs01, Mit97].

KDD wird üblicherweise bezeichnet als der gesamte, nicht triviale Prozess der Identifikation valider, neuartiger, potentiell nützlicher und klar verständlicher Muster in Daten. [FPSS96] In diesem Prozess bezeichnet Data Mining den Teil der Wissensgewinnung, der Analysealgorithmen auf die Daten anwendet, um diese Muster zu entdecken [FPSS96].

Ziel der Projektgruppe ist unter anderem die (Weiter-)Entwicklung einer Mining-Bibliothek, bei der insbesondere temporale Aspekte berücksichtigt werden sollen. Dabei wird das *bitemporal conceptual data model* (BCDM) [JSS94] verwendet. Dieses beinhaltet die Verwaltung von zwei Zeitebenen, der Gültigkeitszeit und der Transaktionszeit.

## 3 Anwendungsszenario

Um die gestellten Ziele zu realisieren, wurde in der Projektgruppe ein Handelszenario analog zu Abbildung 1 aufgebaut. Ein Händler sammelt Daten über seine Kunden, um ihnen ein per-

sonalisiertes Angebot unterbreiten zu können. Ein Problem stellt hierbei die Identifikation des individuellen Kunden bei verschiedenen Einkäufen dar. In jüngster Zeit haben sich Kartensysteme (z.B. Payback [Pay02]) etabliert, die den einzelnen Kunden bei jedem Einkauf identifizieren und so dem Händler ermöglichen, personenbezogene Daten zu speichern.



Abbildung 1: Anwendungsszenario

Ein Händler partizipiert in der Architektur, die wir betrachten, in zweierlei Formen von Handelsbeziehungen. Er handelt einerseits mit Kunden durch den Verkauf von Gütern. Andererseits verkauft er einem Kartenanbieter Informationen und kauft Analyseergebnisse von diesem. Er sammelt, und speichert alle Daten (Verkaufsdaten, Interessenten, Web-logs, etc.), die mit seinem Geschäft zu tun haben. Einen kleinen Teil dieser Daten (die Daten, die mit der angebotenen Karte zu tun haben, i.d.R. die Informationen des Kassensbons) übermittelt er an einen Kartenanbieter. Ihm stehen jedoch weit mehr Informationen zur Verfügung, die er zu Individualisierung seines Angebotes nutzen kann. Hierbei kann er z.T. auf die vom Kartenanbieter gelieferten Analyseergebnisse zurückgreifen. Kennt er die Präferenzen seines Kunden als Individuum, so kann er z.B. in einem elektronischem Shopsystem eine personalisierte Präsentation seiner Waren vornehmen.

Ein Kartenanbieter geht Handelsbeziehungen mit Händlern und Endkunden ein. Um an die Daten zu kommen, die für die Identifikation von Kundengruppen benötigt werden, „bezahlen“ Kartenanbieter Kunden in Form von Prämien (Punktesysteme, Rabatten, u.ä.). An die Daten von Kunden kommen die Kartenanbieter durch die Händler, die ein Entgelt (z.B. in Form von Analyseergebnissen) für die Verarbeitung der Daten erhalten. Der Kartenanbieter sammelt die Daten, integriert und bereinigt sie und speichert sie dauerhaft. Anhand dieser aufbereiteten Daten ist ein Kartenanbieter in der Lage, Analysen verschiedener Art durchzuführen. Hierzu zählen bspw. die Kundengruppierung (vgl. [HK01, ES00]), die Konzeptbeschreibung sowie die Herleitung von Regeln [ES00].

## 4 Realisierung

In der ersten Phase der Projektgruppe wurde ein konkretes Handelsszenario aufgesetzt, das obigen Anforderungen entspricht. Es wurden unabhängig voneinander Datenbankschemata für zwei unterschiedliche Händler sowie einen Kartenanbieter entwickelt. Hierbei wurde besonderer Wert auf die Repräsentation von temporalen Gesichtspunkten der Daten gelegt. In der zweiten Phase wird basierend auf dem Testszenario ein Analysesystem in Form einer Klassenbibliothek entwickelt, mit dessen Hilfe die verteilten Daten integriert, analysiert und ausgewertet werden können. Als Vorbild hierfür dient die Waikato Environment for Knowledge Analysis (WEKA) [WF00], eine in Java implementierte *Machine Learning* Bibliothek. Diese soll metadatengesteuert implementiert werden, um möglichst in beliebigen Szenarien eingesetzt werden zu können. Hierzu soll der bestehende Standard CWM [Obj01] benutzt und eventuell erweitert werden. Der Entwurfsprozeß wird sich dabei am Rational Unified Process (RUP) [Rat02] orientieren. Zur Evaluation der Bibliothek wird ein Online-Shopsystem erstellt.

## Literatur

- [BE01] ANDREAS BAUER und HOLGER GÜNZEL (EDITORS). *Data-Warehouse-Systeme*. Dpunkt.verlag GmbH, **2001**
- [ES00] MARTIN ESTER und JÖRG SANDER. *Knowledge discovery in databases: Techniken und Anwendun gen*. Springer, Berlin, **2000**. ISBN 3-540-67328-8
- [FPSS96] USAMA FAYYAD, GREGORY PIATETSKY-SHAPIRO und PADHRAIC SMYTH. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In EVANGELOS SIMOUDIS, JIA WEI HAN und USAMA FAYYAD, editors, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, Seite 82. AAAI Press, **1996**
- [Heg99] MARKUS HEGLAND. Computational Challenges in Data Mining, **1999**. URL <http://citeseer.nj.nec.com/312469.html>
- [HK01] JIAWEI HAN und MICHELINE KAMBER. *Data Mining, Concepts and Techniques*. **2001**
- [JSS94] C. S. JENSEN, M. D. SOO und R. T. SNODGRASS. Unifying temporal models via a conceptual model. *Information Systems*, 19(7):513–547, **1994**
- [Küs01] ULRICH KÜSTERS. Data Mining Methoden: Einordnung und Überblick. In HAJO HIPPERNER, MATTHIAS MEYER, ULRICH KÜSTERS und KLAUS WILDE, editors, *Handbuch Data Mining im Marketing*. Vieweg, **2001**
- [Mit97] TOM M. MITCHELL. *Machine Learning*. McGraw-Hill, **1997**. ISBN 0-07-042807-7
- [Obj01] OBJECT MANAGEMENT GROUP. Common Warehouse Metamodel (CWM) Specification. Version 1.0, **2001**
- [Pay02] PAYBACK, **2002**. URL <http://www.payback.de/>
- [Rat02] RATIONAL. Rational Unified Process, **2002**. URL <http://www.rational.com/products/rup/index.jsp>
- [WF00] IAN H. WITTEN und EIBE FRANK. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco, **2000**. ISBN 1-55860-552-5