

Seminar - E-Commerce

Machine Learning im Vergleich zu Business Rules im Kontext der Webshop-Personalisierung



- Sommersemester 2013 -

Abgabedatum	27.06.2013
Verfasser	Fabian Hoffmann
Studiengang	Master of Science - E-Commerce
Verwaltungssemester	03
Telefon	(04122) 53815
Adresse	An der Kirche 22 25436 Tornesch
Betreuer (FH Wedel)	Prof. Dr. Holger Schneider
Telefon	(04103) 8048-55
Adresse	Fachhochschule Wedel Feldstraße 143 22880 Wedel

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Abkürzungsverzeichnis	V
1 Einleitung	1
1.1 Gegenstand und Motivation	1
1.2 Zielstellung	2
2 Theorie der Personalisierung	3
2.1 Definition	3
2.2 Elemente der Personalisierung	4
2.3 Klassifikation von Personalisierungstechniken	4
2.3.1 Besucherbasierte Personalisierung	4
2.3.2 Betreiberseitige Personalisierung	5
2.3.3 Automatisierte Personalisierung	6
2.4 Kritik aufgrund der Filter-Wolke	8
3 Theorie der Personalisierungstechniken	9
3.1 Business Rules zur Webshoppersonalisierung	9
3.1.1 Syntaktischer Aufbau von Business Rules	9
3.1.2 Erstellung von Business Rules	10
3.1.3 Verwaltung von Business Rules	11
3.1.4 Auswertung von Business Rules im Webshop-Kontext	12
3.2 Machine Learning zur Webshoppersonalisierung	13
3.2.1 Erstellung eines Machine Learning Systems	14
3.2.2 Verwenden von Machine Learning im Webshop-Kontext	16
4 Vergleich von Business Rules und Machine Learning	17
4.1 Elemente der Personalisierung	17
4.2 Art der Personalisierung	18
4.3 Architektur des Systems	18
4.4 Erstellung von Anwendungsfällen	19
4.4.1 Definition der Logik des Anwendungsfalles	19
4.4.2 Grundsätzliche Denkweise bei der Erstellung von Anwendungsfällen	20
4.4.3 Umgang mit fehlenden Daten	21
4.5 Finden einer optimalen Lösung	22
4.6 Berücksichtigung semantischer Einflussfaktoren	23
4.7 Psychologische Faktoren bei der Verwendung	23

Inhaltsverzeichnis

5 Abschlussbetrachtung	25
Quellenverzeichnis	VI
Eidesstattliche Erklärung	VIII

Abbildungsverzeichnis

2.1	Funktionsweise der besucherbasierten Personalisierung	5
2.2	Funktionsweise der betreiberseitigen Personalisierung	6
2.3	Funktionsweise der automatisierten Personalisierung	7
3.1	Komponenten zur Auswertung von Business Rules	12
3.2	Zusammenspiel von Machine Learning- und Shopsystem	16
4.1	Erzeugung eines Ergebnisses mittels Machine Learning	20
4.2	Finden einer optimalen Lösung	22

Abkürzungsverzeichnis

API	Application Programming Interface
IT	Informationstechnologie
REST	Representational State Transfer

1 Einleitung

1.1 Gegenstand und Motivation

Seit Aufkommen des E-Commerce Ende der neunziger Jahre erfährt dieser ein stetiges Wachstum. In den letzten Jahren wurden in Deutschland Wachstumsraten von über 10 Prozent erzielt - im aktuellen Jahr wird ebenfalls ein solides Wachstum von 12 Prozent erwartet.¹ Andere Länder wie China erreichen sogar Steigerungsraten von über 65 Prozent².

Diese hohen Zuwächse können zur Zeit realisiert werden, da der Markt noch nicht gänzlich gesättigt ist und immer noch eine Nachfragesteigerung seitens der Konsumenten generiert wird. Einige Segmente, wie zum Beispiel der Bekleidungsmarkt, weisen jedoch eine hohe Onlineshop-Dichte auf und es muss zunehmend darauf geachtet werden, dass der Kunde im eigenen Shop kauft und nicht bei der Konkurrenz. Um das zu erreichen, sind Differenzierungsstrategien notwendig, um dem Kunden ein bestmögliches Einkaufserlebnis bieten zu können.

Was bedeutet nun, jedem Besucher ein optimales Einkaufserlebnis zu bieten? Die Antwort gab Jeff Bezos bereits in den späten neunziger Jahren³, als er sagte:

*"If I have 3 million customers on the Web, I should have 3 million stores on the Web."*⁴

Der einzelne Besucher des Webshops soll also damit zufriedengestellt werden, indem ein personalisierter, exakt auf die individuellen Bedürfnisse angepasster Shop präsentiert wird. Zur Umsetzung des Differenzierungsmerkmals werden heutzutage verschiedene Maßnahmen verwendet. Häufig kommen personalisierte Produktempfehlungen, Teaser oder auch eine individuell angepasste Navigation zum Einsatz.

Die Frage ist somit, wie für den einzelnen Shopbesucher entschieden wird, welcher Inhalt optimalerweise angezeigt werden sollte, um die größtmögliche Conversion

¹vgl. Handelsverband Deutschland - HDE e.V. *Der deutsche Einzelhandel*

²vgl. Hagel, *China boomt gewaltig*

³vgl. Nasraoui, *World Wide Web Personalization*

⁴vgl. Drexler, *Icons of Business: Jeff Bezos*, Seite 16

Rate zu erzielen. Bei der technischen Umsetzung dieser Maßnahmen gibt es die verschiedensten Varianten.

1.2 Zielstellung

Ziel dieser Arbeit ist die Beantwortung der Frage, welche technische Realisierung für die Personalisierung von Webshops bestmöglich einsetzbar ist, um für den Besucher dadurch ein optimales - auf ihn zugeschnittenes - Einkaufserlebnis zu ermöglichen. Konkret wird dabei die Umsetzung mittels **Business Rules** sowie automatisierten **Machine Learning** Algorithmen verglichen. Insbesondere soll nicht nur deutlich werden wie diese Techniken funktionieren und wofür sie einsetzbar sind, sondern auch inwieweit sie betriebswirtschaftlich sinnvoll anzuwenden sind.

Nach dem einleitenden **Kapitel 1** wird **Kapitel 2** die theoretische Grundlage zum Thema Personalisierung schaffen.

Kapitel 3 befasst sich mit der theoretischen Funktionsweise von Business Rules sowie von Machine Learning Systemen.

In **Kapitel 4** werden beide vorgestellten Techniken miteinander verglichen. Insbesondere soll dabei die Handhabung aus Sicht des Shopmanagers betrachtet werden, um den Umsetzungsaufwand und damit die Betriebswirtschaftlichkeit zu bewerten. Es wird dazu Bezug auf die zwei vorangegangenen theoretischen Kapitel genommen, die als Grundlage für die Gegenüberstellung dienen.

Im abschließenden **Kapitel 5** wird ein Fazit der Arbeit dargelegt sowie ein Ausblick gegeben.

2 Theorie der Personalisierung

Dieses Kapitel betrachtet Aspekte der Webshop-Personalisierung, die vor allem in Hinblick auf die nachfolgenden Kapitel essentiell sind. Begonnen wird mit einer Definition. Anschließend werden Personalisierungsarten sowie die Klassifizierung von verschiedenen Umsetzungsmöglichkeiten betrachtet. Abgeschlossen wird das Kapitel mit einer kritischen Betrachtung der benutzerindividuellen Gestaltung von Webinhalten.

2.1 Definition

Die Datenmengen im Internet nehmen stetig zu. Bis 2020 wird erwartet, dass das Volumen alle zwei Jahre verdoppelt wird⁵. Da es für Internetnutzer dadurch immer schwieriger wird die passenden Daten in angemessener Zeit zu finden, mussten Mechanismen geschaffen werden, um dies trotzdem zu ermöglichen.

Ein Mechanismus ist die Personalisierung von Webinhalten, welche die Komplexität vorhandener Informationen wieder reduzieren soll. Beim Entfernen von Inhalten gehen diese verloren und sind für den Benutzer nicht mehr einsehbar. Zur Folge hat dies, das der Personalisierungsmechanismus entscheiden muss, welches die für den aktuellen Webseitenbesucher wichtigen Daten sind und dann nur diese zur Anzeige bringt.⁶

Definiert werden kann der Begriff im Kontext des E-Commerce wie folgt:

Personalisierung ist die individuelle Anpassung einer Webseite auf die Bedürfnisse des diese aktuell betrachtenden Besuchers.

Um dies zielgerichtet realisieren zu können, ist Wissen über den Besucher notwendig. Dieses wird in der Regel nicht explizit erfragt, sondern durch Interaktion des Besuchers mit der Webseite erzeugt.⁷

⁵vgl. Handelsblatt, *Wer hebt das Datengold*

⁶vgl. Brudermann, *Webseiten-Personalisierung*, Seite 37f

⁷vgl. Brudermann, *Webseiten-Personalisierung*, Seite 37f

2.2 Elemente der Personalisierung

Auf Webseiten existieren viele Elemente, die sich personalisieren lassen. Die dafür am sinnvollsten erachteten werden in diesem Abschnitt vorgestellt. Es können zwei Hauptkategorien festgelegt werden, denen folgend einzelne Elemente zugeordnet werden⁸:

- Die **Inhaltsebene** bezieht sich auf alle Informationen, die inhaltsbezogen sind. Das bedeutet, diese beeinflussen nur die Semantik der Webseite. Zu dieser Kategorie zählen z.B. Produktvorschläge, Texte, Teaser, Werbung sowie die Sortierung von Produktübersichtsseiten oder Kategorieübersichten. Das wichtigste der aufgezählten Elemente sind die Recommendations von Produkten.
- Die **Präsentationsebene** betrifft alle gestalterischen Segmente der Webseite, die nur das Aussehen auf nicht inhaltlicher Ebene beeinflussen. Nicht inhaltlich bedeutet dabei, dass beispielsweise verschiedene Teaser nicht zu dieser Kategorie gehören, wenn sie an selber Position mit verändertem Inhalt angezeigt werden. Elemente sind unter anderem Farben des Layouts (z.B. Hintergrundfarbe), die grafische Aufbereitung des Menüs oder die Struktur der Gesamtseite. Diese Kategorie der Personalisierung wird zur Zeit nur sehr selten genutzt (z.B. im Rahmen von A/B-Tests, um verschiedene Layouts vor Einführung zu testen⁹).

Im Rahmen dieser Seminararbeit wird in den Beispielen hauptsächlich auf die Inhaltsebene eingegangen. Die Layoutebene lässt sich jedoch mit selben Techniken realisieren.

2.3 Klassifikation von Personalisierungstechniken

Die in den vorigen Abschnitten vorgestellte Personalisierung lässt sich auf viele verschiedene Weisen realisieren. Die für diese Seminararbeit relevanten werden im Folgenden vorgestellt.

2.3.1 Besucherbasierte Personalisierung

Zunächst gibt es die Möglichkeit, dass Besucher einer Webseite die eigenen Interessen selbst bestimmen. Dazu könnte es beispielsweise Formulare geben, die

⁸vgl. Cylogy, Inc. *Personalization Overview*, Seite 38f

⁹Amazon machte dies vor Einführung des aktuellen Layouts

Auswahlmöglichkeiten für präferierte Webinhalte bieten (z.B. Auswahl von Fashionkategorien).¹⁰

Folgende Grafik veranschaulicht das Vorgehen¹¹:

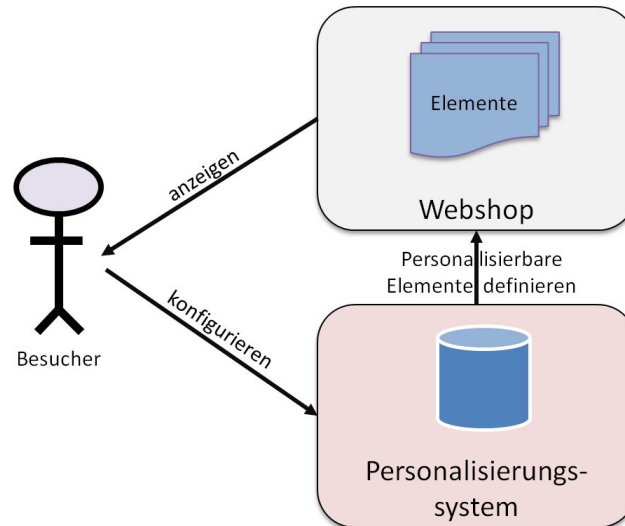


Abbildung 2.1: Funktionsweise der besucherbasierten Personalisierung

Der Besucher (idealerweise registriert, um die Einstellungen zu speichern) konfiguriert das Personalisierungssystem gemäß eigener Präferenzen. Dieses definiert dann die Elemente des Webshos gemäß der Einstellungen. Der Webshop bringt die individuellen Elemente dann für den User zur Anzeige.

Vorteil ist, dass der Besucher komplett selbst entscheiden kann, wie die Webseite gestaltet wird. Nachteile sind, dass manueller Aufwand nötig ist und der Besucher vergessen könnte, dass neue Inhalte verfügbar sind und diese nicht auswählt.¹²

2.3.2 Betreiberseitige Personalisierung

Eine weitere Variante ist die vom Shopbetreiber gesteuerte Individualisierung, bei der festgelegt wird, welche Anpassung der personalisierbaren Elemente bei bestimmten Rahmenbedingungen bzw. Zuständen vorgenommen wird. Das kann zum Beispiel die Festlegung sein, dass zu braunen Schuhen immer ein Teaser mit braunen Gürteln auf dem unteren Teil der Seite angezeigt wird oder das Kunden, die häufig grüne Hosen kaufen, auch grüne Hüte als Vorschläge von dem Empfehlungssystem erhalten. Realisiert wird dies häufig durch regelbasierte Systeme, die mittels gegebener Menge von Regeln die daraus geschlussfolgerten Individua-

¹⁰vgl. Cylogy, Inc. *Personalization Overview*

¹¹vgl. Jochims, *Customer Relationship Management und neue Ansätze*, Kapitel 6

¹²vgl. Cylogy, Inc. *Personalization Overview*

lisierungen durchführen bzw. beim Shop beauftragen.¹³
 Folgende Darstellung zeigt den Ablauf¹⁴:

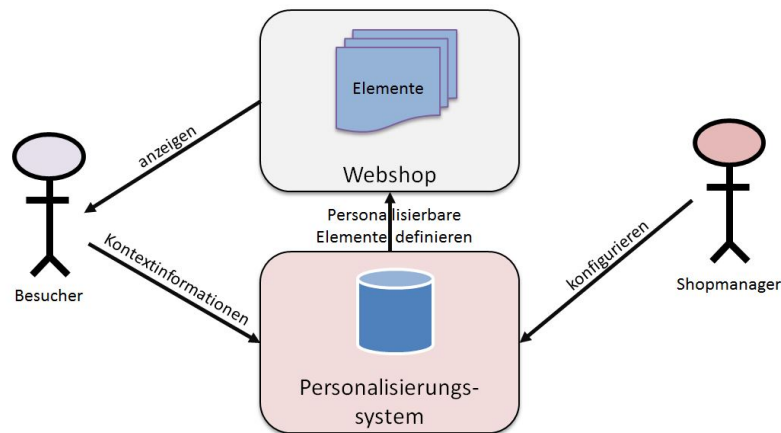


Abbildung 2.2: Funktionsweise der betreiberseitigen Personalisierung

Die Konfiguration wird komplett durch den Shopmanager vorgenommen, der dem Personalisierungssystem dadurch alle notwendigen Informationen liefert, die es zur Bestimmung der userindividuellen Webshop-Elemente benötigt. Der Besucher des Shops liefert ausschließlich Kontextinformation, die zur Bestimmung der eventuell vorhandenen Variablen innerhalb der Konfiguration benötigt werden. Variablen können z.B. der Standort oder die Kaufhistorie des Besuchers sein. Vorteil ist, dass der Shopbetreiber jede Art der Anpassung unter eigener Kontrolle hat und diese dadurch nicht nur alleine bestimmt, sondern auch mit weiteren Marketingkampagnen oder Ähnlichem optimal abstimmen kann. Nachteile sind der erhebliche Pflegeaufwand seitens des Betreibers sowie die eventuell falsch eingeschätzte Sicht des Betreibers, die nicht unbedingt den Wünschen der Besucher entsprechen muss.¹⁵

2.3.3 Automatisierte Personalisierung

Die letzte hier angesprochene Möglichkeit ist die automatische Personalisierung gesteuert durch IT. Es trifft weder der Besucher noch der Shopbetreiber explizit eine Entscheidung über die Personalisierung der Elemente des Webshops. Vom Shopmanager wird lediglich die Hauptaufgabe durch die Wahl eines entsprechend auf den Anwendungsfall passenden Algorithmus festgelegt (siehe Abschnitt 3.2.1). Der Besucher liefert den aktuellen Kontext, in dem er sich befindet (siehe auch Abschnitt 2.3.2).

¹³vgl. Jochims, *Customer Relationship Management und neue Ansätze*, Kapitel 6

¹⁴vgl. Jochims, *Customer Relationship Management und neue Ansätze*, Kapitel 6

¹⁵vgl. Cylogy, Inc. *Personalization Overview*

Nachfolgende Grafik zeigt zunächst das Vorgehen¹⁶:

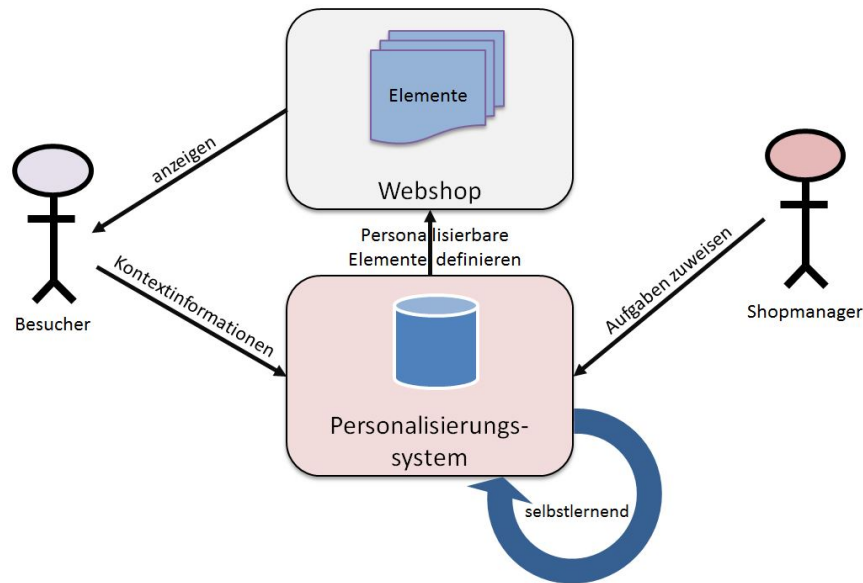


Abbildung 2.3: Funktionsweise der automatisierten Personalisierung

Das Computersystem erzeugt die Anpassung komplett autark.¹⁷ Dafür existieren zwei generalisierte Verfahren, die von den meisten Personalisierungssystemen unterstützt werden¹⁸.

Beim **Collaborative Filtering** werden Besucher der Webseite geclustert, so dass jeder Besucher zu genau einem Cluster gehört. Bei allen sich in einer Gruppe befindenden Personen wird davon ausgegangen, dass sie sich in ähnlichen Situationen ähnlich verhalten und dadurch auch dieselben Präferenzen bezüglich der Webseite (z.B. Produkte) haben. Zur Berechnung von passenden Inhalten oder Layouts werden historische Daten verwendet. Befinden sich z.B. zwei Benutzer *A* und *B* in derselben Profilgruppe und *A* hat Shirt *X* gekauft, dann wird dieses beispielsweise *B* vorgeschlagen, wenn *B* es noch nicht gekauft hat.¹⁹

Das **Content-based Filtering** bezieht sich auf Items, die mit dem Besucher in Verbindung stehen und generiert basierend auf diesen die Ausprägungen personalisierbarer Elemente. Items können dabei im einfachsten Fall Produkte sein, die z.B. gekauft oder angesehen wurden. Weiter betrachtet ist es aber auch möglich, z.B. das Wetter am Standort oder den Standort des Besuchers selbst zu betrachten und darauf basierend Elemente zu individualisieren.²⁰

Vorteil ist die komplett automatische Generierung der auf den Besucher ange-

¹⁶vgl. Jochims, *Customer Relationship Management und neue Ansätze*, Kapitel 6

¹⁷vgl. Jochims, *Customer Relationship Management und neue Ansätze*, Kapitel 6

¹⁸Eigene Recherche bei bekannten Systemen wie z.B. Apache Mahout

¹⁹vgl. Melville und Sindhvani, *Recommender Systems*, Seite 2f

²⁰vgl. Melville und Sindhvani, *Recommender Systems*, Seite 2f

passten Webseite. Ein Nachteil ist, dass der Shopbetreiber komplett auf die Zuverlässigkeit des Systems angewiesen ist und Entscheidungen des Systems aufgrund komplexer Algorithmen häufig nicht gänzlich nachvollziehen kann. Des Weiteren kann es zu Fehlentscheidungen des Systems kommen, da dieses die Semantik des Personalisierungsgegenstandes nur in sehr begrenztem Maße aufnehmen kann (Semantik muss dem System beigebracht werden).

Die drei hier vorgestellten Klassen der Personalisierung müssen nicht immer separat verwendet werden. Es sind auch hybride Lösungen möglich, die verschiedene Techniken miteinander kombinieren. Das macht insbesondere Sinn, wenn für automatisierte Verfahren Ausnahmen gelten sollen. Es es zu empfehlen, dass so wenig Ausnahmen wie möglich erzeugt werden, da sonst der tatsächliche Nutzen des automatisierten Systems infrage gestellt werden sollte. Grund ist, dass das automatische Personalisierungssystem seine Arbeit nur dann optimal erledigen kann, wenn es so wenig Limitierungen wie möglich erhält. Diese Ausnahmen können immer dazu führen, dass die auf Automatisierung ausgelegten Algorithmen nicht mehr effizient funktionieren.

2.4 Kritik aufgrund der Filter-Wolke

An dieser Stelle soll aufgrund der Wichtigkeit die, als eines der im Internet gefährlichsten Phänomene dargestellte, Problematik bei der Erstellung von personalisierten Webseiten betrachtet werden.

Immer mehr Inhalte werden vor der Anzeige für den User gefiltert und dann personalisiert dargestellt. Google wendet dies bei der Erstellung der Suchergebnisse an, Amazon bei der Erstellung von Produktvorschlägen und Retargeting-Firmen bei der Schaltung von Werbebannern. Dies sind nur einige Beispiele. Wenn nun davon ausgegangen wird, dass ein Benutzer eine Anfrage an beliebige Services im Internet stellt und dann perfekt auf seine Interessen angepasste Inhalte geliefert bekommt, dann ist im Vorhinein klar, in welchem Wissenskontext sich der User befindet und auch in Zukunft aufgrund der Personalisierung befinden wird. Der Benutzer hat somit kaum Möglichkeiten seinen Horizont zu erweitern und neue Inhalte zu finden, die bisher nicht auf die Interessen dieses Users passten. Es wird also kontinuierlich gefiltert und der Benutzer hat immer weniger die Chance seinen eigenen Inhaltskontext zu verlassen. Dies wird als Filter-Wolke bezeichnet.²¹

²¹vgl. Hemken, *Filter Bubble*

3 Theorie der Personalisierungstechniken

Das vorliegende Kapitel befasst sich mit den theoretischen Grundlagen von **Business Rules** sowie **Machine Learning** im Kontext der Webshoppersonalisierung. Insbesondere wird auf die Verwendung seitens der Shopmanager eingegangen.

3.1 Business Rules zur Webshoppersonalisierung

Business Rules beziehungsweise Geschäftsregeln sind Regeln der Form *Wenn... dann*, die direkten Einfluss auf das Verhalten bestimmter unternehmerischer Abläufe haben ²². Eine Regel könnte beispielsweise wie folgt definiert sein:

Wenn ein Besucher aus Bayern kommt, **dann** empfehle Wanderschuhe

Die Regel besagt, dass alle Shopbesucher aus Bayern Wanderschuhe empfohlen bekommen.

3.1.1 Syntaktischer Aufbau von Business Rules

Nach dem einführenden Beispiel soll der Begriff Geschäftsregel in diesem Abschnitt definiert werden. Dadurch soll deutlich werden, wie Business Rules syntaktisch gestaltet werden können.

Aufgebaut sind sie durch die Formulierung eines Konditionalsatzes der Form²³:

Wenn A, dann B

Das bedeutet, wenn A erfüllt bzw. wahr ist, dann ist auch B erfüllt bzw. wahr. Der erste Teil des Satzes (A) wird als Prämisse bezeichnet, der zweite Teil des Satzes (B) als Konklusion oder Konsequenz. Es wird von dem Feuern einer Regel gesprochen, wenn A zutrifft und die Regel damit zur Anwendung kommen kann.²⁴

²²vgl. Schacher und Grässle, *Agile Unternehmen*, Seite 17

²³vgl. Kern-Isberner und Beierle, *Methoden Wissensbasierter Systeme*, Seite 72

²⁴vgl. Kern-Isberner und Beierle, *Methoden Wissensbasierter Systeme*, Seite 72

In einer an die Mathematik angelehnten Darstellung kann auch das Implikationsymbol zur Verknüpfung von Prämisse und Konklusion verwendet werden²⁵:

$$\overline{A \rightarrow B}$$

Es ist zulässig, dass die linke Seite der Business Rule aus mehreren Teilprämissen zusammengesetzt ist. Diese werden durch boolesche Operatoren miteinander verknüpft. Am häufigsten werden zur Verkettung OR (\vee), AND (\wedge) und NOT (\neg) benutzt.²⁶ Folgende Definition enthält beispielsweise zwei Teilprämissen, die beide gelten müssen, um die Regel zur Auswertung zu bringen:

$$\overline{\text{Käufer ist Premiumkunde} \wedge \text{Käufer ist Frau} \rightarrow \text{Kostenfreie Duftprobe}}$$

Innerhalb einer Regel existiert immer ein Gerüst, bei dem noch einzusetzende Variablen fehlen. Diese sind durch die Umwelt gegeben und beeinflussen die Auswertung der Regel. Die Variablen werden Fakten genannt.²⁷ Als Beispiel soll folgende Business Rule dienen:

$$\overline{\text{Produkt hat Marge größer 25\%} \rightarrow \text{Wegfall der Versandkosten}}$$

Die Regel ist klar definiert, jedoch lässt sich nicht sagen, ob die Prämisse gilt. Dazu muss die Marge des aktuell betrachteten Produkts bekannt sein. Diese wird der Regel zur Auswertungszeit als Fakt zur Verfügung gestellt.

Nach der Formalisierung des Begriffs wird deutlich, dass Regeln selbst in mathematischer Schreibweise für den Menschen verständlich bleiben, da Konditionalsätze bereits vor vielen tausend Jahren von Menschen als Handlungsanweisungen oder Prognosen benutzt wurden und dadurch intuitiv verständlich sind.²⁸

3.1.2 Erstellung von Business Rules

Wie in den Beispielen gezeigt, bilden Geschäftsregeln reale Business Logik ab. Beim Entwurf von Regeln denken Shopmanager somit bewusst über die Logik des Webshops nach und determinieren diese mittels Regeln. Nicht nur das explizite Nachdenken über die Logik hilft Shopmanagern das eigene Geschäft besser zu verstehen, sondern auch die dadurch automatisch entstehende Dokumentation der Prozesse.

Der Entwurf von Regeln findet somit implizit mit der evolutionären oder aber

²⁵vgl. Kern-Isberner und Beierle, *Methoden Wissensbasierter Systeme*, Seite 72

²⁶vgl. Kern-Isberner und Beierle, *Methoden Wissensbasierter Systeme*, Seite 73ff

²⁷vgl. Wunderlich, *Java Rules Engines*, Seite 17f

²⁸vgl. Kern-Isberner und Beierle, *Methoden Wissensbasierter Systeme*, Seite 73

auch revolutionären Entwicklung des E-Commerce Geschäftsmodells statt, indem die neue Logik unmittelbar mit Hilfe von Geschäftsregeln definiert wird. Bei der Aufstellung von Regeln geht der Ersteller gemäß der syntaktischen Anforderungen, die in Abschnitt 3.1.1 vorgestellt wurden, vor. Jedoch muss insbesondere darauf geachtet werden, dass verschiedene Regeln sich nicht unerwünscht gegenseitig beeinflussen und Reihenfolgebedingungen beim Zutreffen mehrerer Prämissen eventuell einzuhalten sind. Dazu ist es nötig, dass der Shopmanager weiß, wie die Regelbasis ausgewertet wird (siehe Abschnitt 3.1.4).

3.1.3 Verwaltung von Business Rules

Bei der Verwaltung von Regeln ist es wichtig, diese für alle beteiligten Personen zugreif- und verwaltbar zu machen. Dazu sollten alle Business Rules zentral persistiert werden. Um die Abhängigkeit vom eigentlichen Webshop so gering wie möglich zu halten, werden Regeln in einem separaten System abgelegt, das zum Beispiel folgende Funktionen bereitstellt²⁹:

- *Grafische Unterstützung* bei der Erstellung von Regeln. Dies kann beispielsweise die Visualisierung mittels Entscheidungsbäumen sein.
- Eine *Versionsverwaltung* und *Dokumentationsfunktion*, um eine Änderungshistorie zu erhalten. Dadurch können unter anderem ältere Regelbasen betrachtet werden und die Performance zwischen verschiedenen Versionen kann vergleichbar gemacht werden.
- Eine *Benutzerverwaltung*, damit bestimmte User nur speziell definierte Aufgaben im System durchführen können. Z.B. nur Bearbeiten von Regeln, die den Checkout-Prozess betreffen.
- Aufteilung der Regelgesamtheit in Regelteilmengen, um eine *Strukturierung* zu erlangen, die vor allem die Aufteilung der Verwaltungsarbeit zwischen mehreren Shopmanagern erleichtert.

So ein extern vom Webshop existierendes System wird Business Rules Manager oder auch Business Rules Management System (BRMS) genannt.³⁰

Die Abhängigkeit zum Shop sollte aus dem Grund gering gehalten werden, dass der Webshop nicht mit diesen zusätzlichen Aufgaben belastet wird und nur die wirklich notwendigen Informationen vom zentralen Verwaltungssystem erhält.

²⁹vgl. JBoss, Inc. (Red Hat), *Drools Guvnor Dokumentation*

³⁰vgl. JBoss, Inc. (Red Hat), *Drools Guvnor Dokumentation*

Dies sind nur die Regeln, die aktuell vom Webshop verwendet werden sollen. Idealerweise werden die Regeln aus Performancegründen über eine wohldefinierte Schnittstelle an den Webshop übermittelt. Dieser fragt dann nicht bei jeder Regelausführung bei dem BRMS nach den aktuell geltenden Regeln, sondern hat diese sofort in der eigenen Systemumgebung vorliegen. Dies senkt die Fehleranfälligkeit sowie Zugriffsgeschwindigkeit aufgrund der wegfallenden externen Kommunikation mit dem BRMS.

3.1.4 Auswertung von Business Rules im Webshop-Kontext

Die Festlegungen in Form von Regeln wurden bewusst durch eine menschliche Person getroffen und werden seitens der IT nicht infrage gestellt, sondern ohne eigene Analyse lediglich ausgeführt.

Benutzt wird dafür die als Inferenzmaschine bzw. Problemlösungskomponente bezeichnete Einheit, die neben den beiden weiteren Komponenten Regel- sowie Faktenbasis, die als Wissensbasis zusammengefasst werden, existiert³¹. Folgende Grafik zeigt das Zusammenspiel der drei Einheiten:

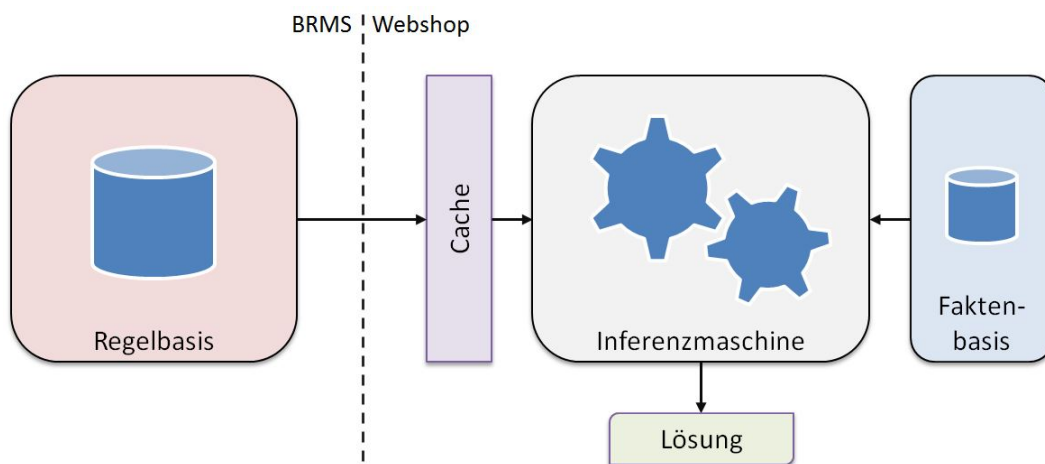


Abbildung 3.1: Komponenten zur Auswertung von Business Rules

Die Regelbasis befindet sich im BRMS und wird dem Webshop über eine Schnittstelle zur Verfügung gestellt. Dieser hält die Regeln in einem Cache oder kleinerem Speicher, der beliebig oft durch das BRMS mit aktuellen Business Rules aktualisiert werden kann (siehe Abschnitt 3.1.3). Die Inferenzmaschine benutzt die sich im Cache befindenden Regeln sowie die Faktenbasis zur Erzeugung eines Ergebnisses. Die Fakten werden aus dem aktuellen Kontext des Webshops erzeugt bzw. sind dort vorhanden.

³¹vgl. Kurbel, *Expertensysteme*, Seite 53;

vgl. Kern-Isberner und Beierle, *Methoden Wissensbasierter Systeme*, Seite 17

Zur Berechnung des Ergebnisses prüft die Inferenzmaschine, welche Prämissen gelten. Die entsprechenden Regeln feuern und erzeugen so neue Fakten, die z.B. definierte Aktionen auslösen können. Zudem dienen neue Fakten als Eingabewerte in weitere Regeln, die dann wiederum zur Ausführung gebracht werden können. Zur Auswertung von Regeln existieren eine Vielzahl Algorithmen, die unter anderem Reihenfolgebedingungen und weitere Aspekte der Ausführung berücksichtigen können. Da diese für die Erstellung von Regeln in dieser Arbeit jedoch nicht relevant sind, findet keine Betrachtung der Berechnungsmethoden statt.³²

3.2 Machine Learning zur Webshoppersonalisierung

Bei der Erstellung von personalisierten Elementen eines Webshops müssen für jeden einzelnen Besucher immer wieder individuelle Entscheidungen getroffen werden. Je besser diese Entscheidungen sind, desto optimaler ist die Internetseite auf den aktuellen Besucher angepasst und entspricht seinen Interessen.

Ziel des Machine Learning ist es, Entscheidungen automatisiert durch den Computer treffen zu lassen und diese stetig automatisch zu verbessern. Machine Learning ist ein Teilgebiet der Künstlichen Intelligenz (KI) und nutzt für die Erstellung von Ergebnissen vorhandene Daten, die bereits gesammelt oder dem System anderweitig zugeführt wurden (z.B. Klick- und Kaufhistorien aller Kunden des Webshops). Innerhalb der vorhandenen Daten wird mittels verschiedenster Algorithmen nach Mustern gesucht, die zu einem Modell generalisiert werden können, um Prognosen über zukünftige Verhaltensweisen zu erzeugen. Beispielsweise könnte ein Algorithmus erkennen, dass Benutzer aus Bayern im September häufig Lederhosen kaufen und diese dann automatisch im September als Produktvorschläge für bayrische Besucher aussteuert. Werden durch diese Aktion keine Mehreinnahmen generiert, erkennt das System dies und erzeugt andere Vorschläge, die auf die entsprechenden Besucher passen.³³

Grundgedanke des Machine Learnings ist die kollektive Intelligenz, die besagt, dass einzelne Individuen einer Gruppe zusammen bessere Ergebnisse erzielen, als würde jede Person der Gruppe alleine agieren. Im Bereich der IT ist damit die Analyse der Verhaltensweisen, Präferenzen oder Ideen der Personen einer Gruppe gemeint, um dadurch neue Erkenntnisse erlangen zu können und insgesamt ein besseres Ergebnis für alle zu schaffen.³⁴

³²vgl. Kurbel, *Expertensysteme*, Seite 53ff

³³vgl. Segaran, *Programming Collective Intelligence*, Seite 3f

³⁴vgl. Segaran, *Programming Collective Intelligence*, Seite 2f

3.2.1 Erstellung eines Machine Learning Systems

Bei der Verwendung eines Machine Learning Systems steht nicht die Definition von Richtlinien zur Erstellung der Ergebnisse im Vordergrund, sondern die Auswahl des richtigen Algorithmus, der die Erzeugung von Resultaten automatisiert durchführt. Shopmanager müssen sich somit zunächst mit der Theorie der Algorithmen befassen, um dann einen oder mehrere passende für den umzusetzenden Anwendungsfall zu bestimmen.

Im Folgenden werden drei der wichtigsten in der Webshop-Personalisierung eingesetzten Vorgehensweisen (Recommendation Engines, Classification sowie Clustering) beschrieben³⁵. Die Betrachtung einzelner Implementierungen der Algorithmen findet hier nicht statt, da dies im Rahmen dieser Seminararbeit nicht von Bedeutung ist.

- Täglich haben Menschen extrem viele Entscheidungen zu treffen, da es unzählige Möglichkeiten gibt, aus denen ausgewählt werden kann. Die Produktvielfalt sowie die Anzahl der verfügbaren Informationen nimmt täglich zu (siehe auch Abschnitt 1.1). Sei es z.B. die Entscheidung für einen Film aus einer Auswahl von knapp einer halben Millionen Filmen in einem Onlineshop³⁶. Eine erste Möglichkeit des Onlineshops zur Unterstützung der Entscheidungsfindung wäre die Schaltung von Werbeflächen, die für alle Besucher dieselben Filme vorschlagen. Jedoch kann diese Werbung höchstens das Durchschnittsinteresse treffen, jedoch nie die Aufmerksamkeit aller erlangen. Eine Verbesserung sind Vorschläge für jede einzelne Person, die den Interessen dieser entsprechen (z.B. anhand getätigter Käufe).³⁷

Recommendation Engines sind die am häufigsten eingesetzte Personalisierungsart und haben genau das als Ziel: Bestmögliche und damit personalisierte Vorschläge für jede einzelne Person zu erzeugen. Im Kontext von Webshops werden insbesondere Produkte vorgeschlagen, um dadurch die Aufmerksamkeit auf diese zu lenken und dem Besucher, bereits vor der eigenen Suche nach Artikeln im Shop, auf eine interessante Artikeldetailseite umzuleiten. Idealerweise besteht dann bereits ein Kaufinteresse. Auf der Artikeldetailseite könnten ergänzende Recommendations angezeigt werden, die Zubehör oder weitere zu dem betrachteten Artikel passende Produkte bekanntmachen.³⁸

³⁵vgl. Owen u. a., *Mahout in Action*, Seite 3f

³⁶Eigene Recherche auf <http://www.amazon.de>

³⁷vgl. Marmanis und Babenko, *Intelligent Web*, Seite 69f

³⁸vgl. Owen u. a., *Mahout in Action*, Seite 3;
vgl. Segaran, *Programming Collective Intelligence*, Seite 7

- **Clustering**-Techniken bilden aus einer großen Menge von beliebigen Objekten Gruppen bzw. Cluster, wobei die Elemente innerhalb eines Clusters in irgendeiner Weise Ähnlichkeiten aufweisen (z.B. ähnliche Eigenschaften). Dadurch wird es möglich, Strukturen oder Hierarchien festzustellen, die im gesamten Datensatz sonst nicht ohne weiteres zu erkennen wären. Google News z.B. kategorisiert die News-Beiträge mittels Clustering-Methoden.³⁹ Im Kontext von Webshops können z.B. Kunden mit ähnlichen Interessen gruppiert werden, um gezielte Marketing-Aktionen durchzuführen. Einer Gruppe, die gemäß Kaufverhalten an Actionfilmen interessiert ist, könnte dann z.B. ein Angebot für eine vorzeitige Bestellung eines bald erscheinenden Actionfilms gemacht werden.
- Zur **Classification** gehörende Algorithmen entscheiden, in welche bereits bestehende Gruppe bzw. Klasse ein bisher nicht zugeordnetes Objekt gehört. Die Klassen müssen dabei im Vorhinein bekannt sein. Durch die Klassifizierung ist es möglich, neue Objekte direkt in einen Kontext einzuordnen und sie dadurch schnellstmöglich zu verstehen, da sie in irgendeiner Weise Ähnlichkeiten zu den anderen Objekten derselben Klasse aufweisen.⁴⁰ Clustering und Classification haben somit sehr ähnliche Zwecke - die Strukturierung von Elementen. In Webshops ist eine Kombination beider Techniken möglich, wie folgendes Vorgehen verdeutlichen soll: Zunächst werden alle Kunden einem Clustering unterzogen, damit Gruppen entstehen. Jeder neue Kunde wird dann mittels Classification in eine existierende Gruppe eingeordnet, damit nicht bei jedem neuen Kunden ein komplett neues Clustering durchgeführt werden muss und dadurch Performancevorteile erlangt werden. Die Entscheidung, in welche Klasse ein Neukunde gehört, muss unter Umständen in nur wenigen Sekunden getroffen werden, um beispielsweise die Kaufbestätigungsmail gemäß Gruppenzugehörigkeit zu individualisieren. Das Clustering sollte dann in wohldefinierten Zeitabständen erneut durchgeführt werden (z.B. jede Nacht oder einmal wöchentlich).

Jede der drei Techniken kann mit einer Vielzahl verschiedener Algorithmen umgesetzt werden. Es ist nicht immer klar, welches der geeignetste für den jeweiligen Anwendungsfall ist. Aus dem Grund ist häufig ein Testen mehrerer Implementierungen nötig.

³⁹vgl. Owen u. a., *Mahout in Action*, Seite 3f;

vgl. Marmanis und Babenko, *Intelligent Web*, Seite 121ff

⁴⁰vgl. Marmanis und Babenko, *Intelligent Web*, 164ff

vgl. Owen u. a., *Mahout in Action*, Seite 4f;

3.2.2 Verwenden von Machine Learning im Webshop-Kontext

Machine Learning Techniken greifen auf eine Vielzahl von Daten gleichzeitig zu (z.B. beim Clustering auf alle zu clusternden Elemente). Deshalb ist es sinnvoll, eine direkte Schnittstelle zur Datenhaltung des Webshops zu haben, jedoch die Performance des Shopsystems aufgrund der teilweise sehr komplexen und umfassenden Algorithmen (z.B. Durchführung des Clusterings) nicht mit zusätzlichen Aufgaben zu belasten.

Um die Unabhängigkeit zu wahren und die möglichen Fehlerquellen (z.B. zu hohe Last des Server- oder Datenbanksystems) zu reduzieren, sollte das Machine Learning System somit separat vom Webshop als Applikation bereitgestellt werden. Vor allem, da Machine Learning Systeme sich in der Regel fortlaufend weiterentwickeln und dadurch stetig Berechnungen durchführen. Der Zugriff des Shops erfolgt über eine wohldefinierte API (z.B. REST-API). Für den Fall, dass das externe System nicht verfügbar ist, muss es default-Werte im Shopsystem geben.

Folgende Darstellung verdeutlicht das Zusammenspiel beider Applikationen:

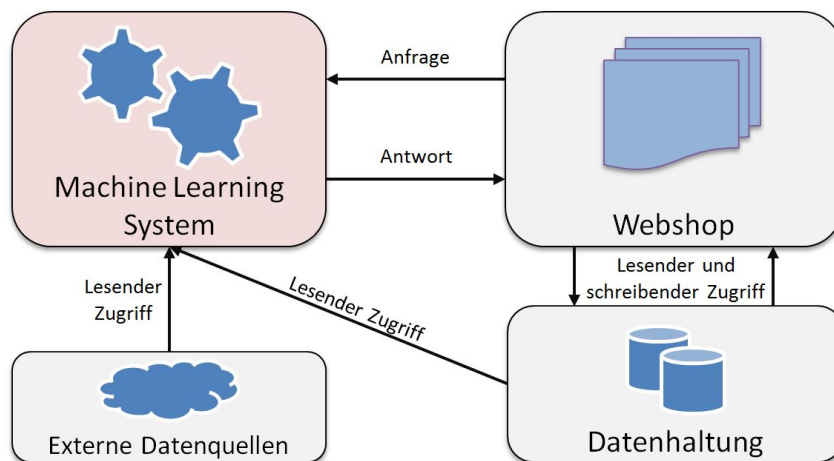


Abbildung 3.2: Zusammenspiel von Machine Learning- und Shopsystem

Das Machine Learning System erhält Anfragen vom Webshop und liefert dazu passende Ergebnisse. Für die Berechnung kann lesend auf Daten des Webshops zugegriffen werden. Zudem können externe Datenquellen (z.B. Wetterinformationen) verwendet werden.

4 Vergleich von Business Rules und Machine Learning

Dieses Kapitel soll die beiden vorgestellten Techniken zur Umsetzung von Personalisierung miteinander vergleichen sowie dabei Vor- und Nachteile bei der Verwendung aufzeigen.

Begonnen wird mit einer Verwendbarkeitsanalyse bezüglich der möglichen zu personalisierenden Elemente (siehe Abschnitt 2.2) eines Webshops. Anschließend werden die zwei Vorgehensweisen gemäß der in Abschnitt 2.3 vorgestellten Personalisierungstechniken klassifiziert sowie die Architektur der Systeme betrachtet. Den Hauptteil des Kapitels bildet Abschnitt 4.4, der die konkrete Erstellung von Anwendungsfällen analysiert und dabei auch den zeitlichen Aufwand betrachtet. Die weiteren Abschnitte werden das Finden einer optimalen Lösung, die Berücksichtigung semantischer Faktoren sowie psychologische Rahmenbedingungen bei der Verwendung der verschiedenen Systeme betrachten.

4.1 Elemente der Personalisierung

In Bezug auf die Segmente einer Webseite (siehe Abschnitt 2.2), die sich personalisieren lassen, werden sowohl mit Business Rules als auch mit Machine Learning Techniken alle Elemente unterstützt. Es gibt jedoch Anwendungsfälle, die sich eher für die eine oder andere Technik eignen. Sollen nur statische Webseiten-Elemente mit geringer Komplexität angepasst werden, dann können Business Rules die Aufgabe häufig schneller lösen, als wenn ein Machine Learning System aufgebaut wird. Dabei kann es beispielsweise um die Schaltung von Werbung über Nachtmode nach 20:00 Uhr gehen. Handelt es sich jedoch um die Erstellung von Empfehlungen, müsste der Shopmanager sehr viele Regeln definieren, damit das System personalisiert genug arbeitet und jedem Besucher unterschiedliche, auf ihn abgestimmte, Ergebnisse anzeigen kann. Hier wäre der Aufwand mittels Business Rules sehr hoch und es könnte sich ein Machine Learning System empfehlen.

4.2 Art der Personalisierung

In Abschnitt 2.3 wurden verschiedene Klassen von Personalisierungsarten aufgezeigt. Die Techniken Business Rules sowie Machine Learning werden in diesem Abschnitt mit Hilfe dieser klassifiziert.

Business Rules werden im Bereich der *betreiberseitigen Personalisierung* (siehe Abschnitt 2.3.2) eingesetzt. Der Betreiber eines Webshops entscheidet darüber, wie individuelle Ergebnisse für die einzelnen Besucher erzeugt werden. Der Besucher hat keine Chance (außer mit seinen Eigenschaften selbst, die als Input für die Business Rules gelten) die Regeln zu beeinflussen. Ebenso werden die Regeln nicht automatisiert verändert.

Machine Learning Methoden hingegen sind in die Klasse der *automatisierten Personalisierung* (siehe Abschnitt 2.3.3) einzuordnen. Der Betreiber bestimmt zwar die Art der Eingabeparameter sowie die Art des Outputs (siehe Abschnitt 4.4), die Analyse selbst jedoch passt sich automatisch an neue Gegebenheiten an und muss nicht manuell konfiguriert werden. Der Besucher hat ebenfalls keine Chance auf die Auswertung selbst einzuwirken (bis auf durch die naturgegebenen Eigenschaften des Besuchers als Input).

4.3 Architektur des Systems

Hinsichtlich der Systemarchitektur unterscheiden sich beide betrachteten Techniken, insbesondere aufgrund der unterschiedlichen Dynamik, voneinander.

Die **Geschäftsregeln** selbst werden idealerweise innerhalb des Webshopssystems in den Quellcode eincompiliert, damit eine möglichst hohe Performance bei der Auswertung erzielt werden kann (siehe Abschnitt 3.1.4)⁴¹. Möglich ist dies, da die Regeln sich nicht fortlaufend ändern, sondern nur, wenn der Shopmanager explizit Anpassungen vornimmt. Nach der Anpassung müssen diese zu einem fest definierten Zeitpunkt kontrolliert in die Codebasis des Webshops überführt werden. Die Administration der Business Rules findet jedoch außerhalb des Shops statt, um den Webshop nicht mit Zusatzaufgaben zu belasten.

Machine Learning Systeme sollten extern vom Webshop arbeiten und auch dort die Ergebnisse produzieren, da fortlaufend und automatisiert Änderungen an der Auswertungslogik vorgenommen werden (siehe Abschnitt 3.2.2). Der Webshop würde ansonsten einer zu hohen Dynamik innerhalb der Logik ausgesetzt sein und fehleranfälliger werden.

⁴¹vgl. JBoss, Inc. (Red Hat), *Drools Expert Dokumentation*

4.4 Erstellung von Anwendungsfällen

Dieser Abschnitt zeigt, wie Anwendungsfälle mit beiden vorgestellten Techniken erstellt werden. Nachdem die grundsätzliche Vorgehensweise der Logikerstellung betrachtet wurde, wird in einem zweiten Abschnitt die gänzlich verschiedene Denkweise bei der Erstellung herausgestellt. Abschließend wird auf die Problematik fehlender Daten, gerade zum Release-Zeitpunkt des Anwendungsfalles, eingegangen.

4.4.1 Definition der Logik des Anwendungsfalles

Ein auf **Business Rules** basierender Anwendungsfall wird definiert, indem eine Regelbasis erstellt wird. Diese kann je nach Komplexität des use cases mehrere tausend Regeln umfassen. Das bedeutet, Shopmanager machen sich konzeptionelle Gedanken über alle Situationen, die eintreten können und formulieren diese in Form von Regeln. Es kann somit ein zeitlich sehr hoher Aufwand entstehen. Zudem besteht die Gefahr, dass Regeln vergessen, keine default-Rules erstellt (z.B. nicht bei jedem Besucher ist der Wohnort bekannt) oder die Reihenfolge der Regelauswertung nicht berücksichtigt wird. Das Reihenfolgeproblem tritt insbesondere bei großen Regelbasen auf, da der Überblick sehr schnell verloren gehen kann. Des Weiteren tritt hoher Aufwand auf, wenn sich die Eigenschaften der Besucher stark verändern oder neue Produkte bzw. Kategorien hinzugefügt werden, die bisher nicht berücksichtigt wurden. Diese müssen manuell so schnell wie möglich nachgepflegt werden, um die optimal zu erzielende Conversion Rate des Shops durch default-Regeln (z.B. default-Produktvorschläge) nicht zu senken. Sollen beispielsweise Produktempfehlungen angesteuert werden, muss für jeden Artikel exakt festgelegt werden, welcher andere Artikel dazu passt. Alternativ könnte dies auf Kategorieebene definiert werden (welche Kategorie passt zu welcher Kategorie), jedoch wäre das Ergebnis dann ungenauer. Sollen Besucherpräferenzen berücksichtigt werden, muss mittels Regeln festgelegt werden, welche Besuchereigenschaften für bestimmte Produkte erfüllt sein müssen, damit sie gemocht werden. Auch dies würde einen erheblichen Aufwand erzeugen.

Bei der Umsetzung eines **Machine Learning** Systems werden keine Regeln definiert, die das exakte Verhalten des Systems bestimmen und jede mögliche Situation berücksichtigen müssen. Hier wird der Betrachtungskontext festgelegt, den das System optimieren soll. Dazu gehört der Input für die Analyse der Eingabeparameter, ein Algorithmus zur Erzeugung des Zwischenergebnisses sowie der

Output, der aus dem Analyseergebnis erzeugt werden soll:

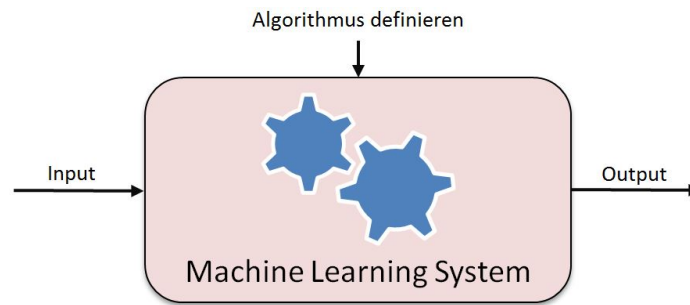


Abbildung 4.1: Erzeugung eines Ergebnisses mittels Machine Learning

Die zu verwendenden *Eingabeparameter* werden durch den Shopmanager festgelegt. Beispielsweise Kunden- und Produktdatenbank. Zur Berechnung muss der für den Anwendungsfall am besten geeignete *Algorithmus* bestimmt werden (manuell oder durch multivariate Tests). Der Algorithmus könnte als Zwischenergebnis z.B. Benutzer gemäß ihrer Produktpräferenzen clustern. Die Ähnlichkeit wird beispielsweise anhand der Käufe sowie Klicks auf Produkte berechnet. Die Produkte werden anhand ihrer spezifischen Attribute zu verschiedenen Benutzergruppen zugeordnet (auch Teil des Zwischenergebnisses). Durch die Gruppenzugehörigkeit jedes Benutzers kann diesem immer ein für ihn in der Regel interessantes Produkt vorgeschlagen werden. Das wäre dann das *Ergebnis* des Systems. Bei Veränderungen der Kundenbasis oder der Produkte wird dies automatisch vom System berücksichtigt. Das heißt, die Kundensegmente können sich jederzeit ändern, ohne dass das System durch den Shopmanager angepasst werden muss. Noch nicht bekannte Besucher können ebenfalls berücksichtigt werden, indem eine Klassifizierung (siehe Abschnitt 3.2.1) stattfindet.

Mit der Anzeige personalisierter Teaser kann sehr ähnlich zu der Empfehlung von Produkten verfahren werden: Jeder Teaser wird bestimmten Produkten zugeordnet und dadurch für das Machine Learning System automatisch auch anderen sehr ähnlichen Produkten zugewiesen. Der Output wären somit keine Produkte, sondern ein oder mehrere Teaser, die dem User angezeigt werden.

4.4.2 Grundsätzliche Denkweise bei der Erstellung von Anwendungsfällen

Es lässt sich herausstellen, dass für beide Techniken eine gänzlich unterschiedliche Denkweise besteht.

Beim **Business Rules** Konzept werden konkrete Eingabewerte definiert, die

dann direkt bestimmen, was als Ausgabe erzeugt wird. Beim **Machine Learning** wird nur die Art der Eingabeparameter bestimmt. Was aber tatsächlich als Ausgabeergebnis erscheint, wird vom System selbst erlernt. Bei der Business Rules-Technik würde z.B. eingegeben werden, dass bei einem bestimmten Wetter bestimmte Produkte vorgeschlagen werden:

Wenn die Sonne scheint, **dann** empfehle die 5 beliebtesten Sonnenbrillen
.. weitere Regeln für jede Wetterlage ..

Ein Machine Learning System jedoch würde nur wissen, dass das Wetter als Eingabeparameter dient und die Ausgabe die fünf beliebtesten Produkte einer Kategorie sind. Es würde nun im Laufe der Zeit selbst lernen, dass bei Sonne am häufigsten Sonnenbrillen gekauft werden und dann diese Kategorie als Output berechnen. Ändert sich das Kaufverhalten bei Sonne, wird das System dies im Gegensatz zu dem Business Rules System automatisch erlernen.

4.4.3 Umgang mit fehlenden Daten

Wenn ein neues Personalisierungssystem eingeführt wird, benötigt dieses Daten zur Erzeugung von Ergebnissen. Wenn diese jedoch zum Launch eines Anwendungsfalles nicht vorliegen und erst im Verlauf der Zeit gesammelt werden müssen, wird von dem sogenannten Kaltstart-Problem gesprochen⁴².

Business Rules können häufig auch ohne große Datensammlungen arbeiten, da die Eingabedaten nicht auf großflächigen Analysen der Gesamtdaten mittels Data Mining oder Ähnlichem basieren, sondern exakt definierte Eingabeparameter erwarten, die der aktuelle Besucherkontext automatisch zur Verfügung stellt. Beispielsweise könnten Regeln definiert sein, dass Besucher mit bestimmten Produktkategorien im Warenkorb bestimmte andere Produkte als Zusatzprodukte angeboten bekommen. Dazu sind vorab keine weiteren Daten nötig.

Machine Learning Techniken hingegen basieren auf der Analyse von großen Datenmengen, um dort Muster zu erkennen und daraus Generalisierungen zu erstellen. Ohne ein gewisses Datenvolumen können diese Analysen nicht durchgeführt werden und es kann kein individuelles Ergebnis erzeugt werden, da schlicht das Modell zur Bestimmung fehlt. In diesem Fall muss auf default-Werte zurückgegriffen werden, die z.B. mittels Business Rules zu definieren sind. Sobald genügend Informationen für weiterführende Analysemethoden vorhanden sind, werden die default-Regeln automatisch durch das System ausgeschaltet. Maschinell lernende Systeme sind somit im Gegensatz zu Business Rules sehr stark durch die

⁴²vgl. Schneider, *E-Commerce Systeme*, Kapitel 5.3

Kaltstart-Problematik betroffen.

4.5 Finden einer optimalen Lösung

Zunächst stellt sich die Frage, was eine optimale Lösung ist. Im Bereich der personalisierbaren Webshop-Elemente ist die Aufgabe dabei, genau die Ausprägung des Elementes auszusteuern, die dem aktuell betrachteten Besucher des Shops den größten Mehrwert bietet. Der Webshop soll also mit den Möglichkeiten (z.B. verfügbare Daten), die dem Berechnungsalgorithmus zur Verfügung stehen, am besten personalisiert werden.

Folgende Grafik visualisiert den Sachverhalt⁴³:

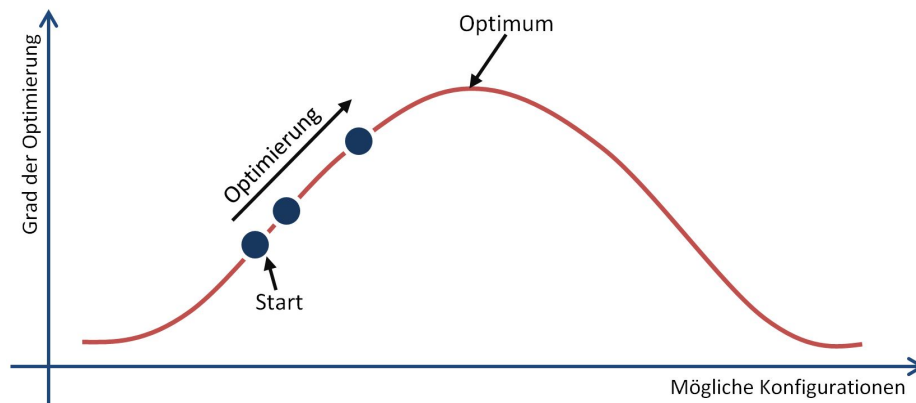


Abbildung 4.2: Finden einer optimalen Lösung

Business Rules werden manuell angelegt und modifiziert. Das heißt, zunächst legt der Shopmanager Business Rules an, von denen er glaubt, dass sie in Bezug auf die Individualisierung am besten performen (Punkt *Start* in Abbildung 4.2). Wird im laufenden Betrieb festgestellt, dass die Regeln noch besser formuliert werden können, wird der Shopmanager die Regeln wieder umschreiben und dem Shop eine aktualisierte Version zur Verfügung stellen (Pfeil *Optimierung* in Abbildung 4.2). Ob eine Verbesserung erzielt wird, muss erst getestet werden. Jeder Veränderungsschritt ist somit ein manueller Eingriff, bei dem fast nie sicher ist, ob tatsächlich ein besseres Ergebnis erzielt wird und dadurch das *Optimum* erreicht wird. Selbst wenn das optimale Ergebnis nach mehreren Iterationen erzielt wird, unterliegt dieser Punkt selbst einer Eigendynamik, die nur schwer abzuschätzen ist. Der optimale Punkt selbst kann wandern und dadurch eine erneute Veränderung der Business Rules erfordern. Zieht ein Besucher des Webshops beispielsweise von

⁴³vgl. WibiData, *Big Data Workshop*

Hamburg nach Bayern, erfordert dies eventuell einen anderen Betrachtungskontext und somit andere Regeln, um ein besseres Ergebnis zu erzielen.

Machine Learning Algorithmen hingegen erfordern keine manuelle Anpassung von den Shopmanagern, sondern optimieren sich fortlaufend selbst. Verändert sich der Umgebungskontext oder ist das Ergebnis nicht zufriedenstellend, wird der Algorithmus dies merken und die Berechnungsmethode automatisiert verändern. Das Optimum kann dadurch nicht nur schneller erreicht werden, sondern das System reagiert auch bei der Verschiebung des Optimums in Richtung der x-Achse vollständig autark ohne menschlichen Eingriff.

4.6 Berücksichtigung semantischer Einflussfaktoren

Häufig ist die Betrachtung von Informationen notwendig, welche über die in Datenbanken direkt auswertbaren hinausgehen und von Computern semantisch nicht unmittelbar erfasst werden können. Dabei kann es sich beispielsweise um kulturelle Gegebenheiten verschiedener geographischer Regionen handeln, die ein Computer nicht wie Produktdaten oder Ähnliches aus den Daten herauslesen kann. Ein Mensch hingegen kann auch externe Informationen und Bezugskontexte nutzen und in die Erstellung von Anwendungsfällen einbeziehen.⁴⁴

Machine Learning steht somit den **Business Rules** nach, sobald Informationen benötigt werden, die semantische Bewertungen benötigen. Soll diese maschinell erfolgen, muss dem Computersystem zunächst eine entsprechende Datenbasis geschaffen werden, die fortlaufend aktuell gehalten werden muss, wenn sich die einbezogenen semantischen Informationen verändern.

4.7 Psychologische Faktoren bei der Verwendung

Nachdem die vorigen Abschnitte eher technische Aspekte betrachteten, wird dieser Abschnitt sich mit den psychologischen Verhaltensweisen der Menschen beschäftigen, die die zwei analysierten Techniken verwenden.

Bei der Nutzung von **Business Rules** ist dem Shopmanager exakt bekannt, wie sich die Logik des Systems verhält, da er selbst die Regeln mit dazugehörigen Inputs und Outputs definiert hat. Jedenfalls, wenn die Theorie der Regelauswertung beherrscht wird. Angenommen, dass die Technik fehlerfrei funktioniert, muss die das System bedienende Person nur sich selbst vertrauen.

Anders ist es bei der Verwendung von **Machine Learning** Algorithmen. Hier

⁴⁴vgl. Segaran, *Programming Collective Intelligence*, Seite 4

muss dem Computer vertraut werden, da lediglich die Vorgehensweise der Maschine bestimmt wird, jedoch nicht exakt klar ist, was diese daraus macht. Auch die Nachvollziehbarkeit ist nicht bei allen Algorithmen, wie zum Beispiel bei neuronalen Netzen, gegeben. Der Mensch muss dem System somit vertrauen und definiert nur noch die Rahmenbedingungen, in denen es arbeitet.

Des Weiteren könnten Shopmanager denken, dass ihre Arbeitskraft bald nicht mehr benötigt wird, wenn alle Optimierungen, die sonst manuell durchgeführt wurden, nun durch Computer getätigt werden. Jedoch ist dieser Gedankengang nicht begründet, da der Shopmanager seine Fähigkeiten nun viel sinnvoller einsetzen kann. Zum Beispiel in den Entwurf von noch viel komplizierteren Konzepten, die mit den Möglichkeiten des Machine Learnings umgesetzt werden können und bei denen regelbasierte Systeme an ihre Grenzen stoßen würden.

5 Abschlussbetrachtung

Die vorliegende Arbeit hat die zur Personalisierung von Webshops einsetzbaren Techniken Business Rules sowie Machine Learning in Bezug auf Verwendbarkeit und Betriebswirtschaftlichkeit miteinander verglichen.

In **Kapitel 2** zeigte sich, dass die Personalisierung aufgrund der zunehmenden Datenmengen im E-Commerce einen immer höheren Stellenwert einnimmt und dadurch effiziente Basistechnologien vorhanden sein müssen. Insbesondere die automatisierte Personalisierung ist sehr zukunftssträchtig, da optimale Ergebnisse mit möglichst geringem zeitlichen Aufwand seitens der Shopmanager im Idealfall kombiniert werden.

Kapitel 3 hat die technische Realisierung der beiden Umsetzungsmöglichkeiten beschrieben, um diese nachfolgend in **Kapitel 4** zu vergleichen. Bei den Business Rules handelt es sich um eine vom Betreiber gesteuerte Personalisierung, die sich im Gegensatz zum Machine Learning nicht automatisiert weiterentwickelt. Dadurch entsteht betriebswirtschaftlich betrachtet ein erheblicher Unterschied in Bezug auf die benötigten humanen Ressourcen. Um eine optimale Lösung zu finden sind häufig viele Iterationen und eine fortlaufende Kontrolle notwendig, die beim Business Rules Ansatz komplett manuell durchgeführt werden müssen. Machine Learning Systeme hingegen kontrollieren und optimieren sich selbst. Eine menschliche Kontrolle sollte in jedem Fall trotzdem durchgeführt werden. Sollen komplexe semantische Einflussfaktoren berücksichtigt werden, kann die Verwendung von Business Rules aufgrund der schlechten Auswertungsmöglichkeiten von Bedeutung seitens der IT vorteilhaft sein. Ebenso, wenn Anwendungsfälle mit nur sehr geringer Dynamik umgesetzt werden sollen.

Um die Vorteile beider Techniken bestmöglich zu nutzen, kann es sinnvoll sein, einen hybriden Ansatz zu implementieren. Machine Learning wird für komplexe und sich sehr dynamisch verändernde Anwendungsfälle benutzt. Ausnahmen innerhalb der komplexen Anwendungsfälle sowie eher statische Anwendungsfälle werden mittels Business Rules umgesetzt. Ausnahmen sollten dabei so wenig wie möglich definiert werden, um die Effizienz der automatisierten Algorithmen nicht zu verringern.

Quellenverzeichnis

Literaturverzeichnis

- Brudermann, Thomas. *[Webseiten-Personalisierung] für anonyme Besucher*. Nordestedt: GRIN Verlag, 2007.
- Drexler, Kateri. *[Icons of Business: Jeff Bezos]*. Westport, CT (US): Greenwood Publishing Group, 2007.
- Kern-Isberner, Gabriele und Christoph Beierle. *[Methoden Wissensbasierter Systeme] - Grundlagen, Algorithmen, Anwendungen*. 4. Aufl. Wiesbaden: Vieweg + Teubner, 2008.
- Kurbel, Karl. *Entwicklung und Einsatz von [Expertensysteme]n - Eine anwendungsorientierte Einführung in wissensbasierte Systeme*. 2. Berlin, Heidelberg: Springer-Verlag, 1989, 1992.
- Marmanis, Haralambos und Dmitry Babenko. *Algorithms of the [Intelligent Web]*. Greenwich (US): Manning Publications Co., 2009.
- Owen, Sean u. a. *[Mahout in Action]*. New York (US): Manning Publications Co., 2012.
- Schacher, Markus und Patrick Grässle. *[Agile Unternehmen] durch Business Rules - Der Business Rules Ansatz*. Berlin, Heidelberg: Springer-Verlag, 2006.
- Segaran, Toby. *[Programming Collective Intelligence]*. Sebastopol (US): O'Reilly Media, Inc., 2007.
- Wunderlich, Lars. *[Java Rules Engines] - Entwicklung von regelbasierten Systemen*. Frankfurt: entwickler.press, 2006.

Quellen im Internet

- Cylogy, Inc. *[Personalization Overview]*. San Francisco (US): http://www.cylogy.com/library/personalization_overview-kb.pdf, Zugriff am 12.06.2013.
- Hagel, Philipp. *[China boomt gewaltig]: E-Commerce Markt wächst um 66.5getrieben von Mobile und Social Media*. Butzbach: <http://www.ecommerce-lounge.de/china-ecommerce-waechst-10675>, Zugriff am 07.06.2013.

- Handelsblatt. [*Wer hebt das Datengold*]. Düsseldorf: <http://www.handelsblatt.com/unternehmen/digitale-revolution-der-wirtschaft/big-data-wer-hebt-das-datengold-seite-all/7613108-all.html>, Zugriff am 12.06.2013.
- Handelsverband Deutschland - HDE e.V. [*Der deutsche Einzelhandel*]. Berlin: http://www.einzelhandel.de/index.php/presse/zahlenfaktengrafiken/item/download/5133_e6af50605ac9198aaab3b0a1da81fd2f.html, Zugriff am 08.06.2013.
- Hemken, Heiner. [*Filter Bubble*]: *Folgen personalisierter Informationen*. Norden: <http://www.seo-united.de/blog/internet/filter-bubble-folgen-personalisierter-informationen.htm>, Zugriff am 12.06.2013.
- JBoss, Inc. (Red Hat). [*Drools Expert Dokumentation*]. Atlanta (US): http://docs.jboss.org/drools/release/6.0.0.Beta3/drools-expert-docs/html_single/index.html, Zugriff am 16.06.2013.
- [*Drools Guvnor Dokumentation*]. Atlanta (US): http://docs.jboss.org/drools/release/6.0.0.Beta2/drools-guvnor-docs/html_single/index.html, Zugriff am 02.06.2013.
- Melville, Prem und Vikas Sindhwani. [*Recommender Systems*]. New York (US): <http://www.vikas.sindhwani.org/recommender.pdf>, Zugriff am 12.06.2013.
- Nasraoui, Olfa. [*World Wide Web Personalization*]. Louisville (US): http://webmining.spd.louisville.edu/Websites/PAPERS/book_chapter/FINAL-Nasraoui-WWW-Personalization.htm, Zugriff am 07.06.2013.

Inhalte aus Vorlesungen

- Jochims, Dr. Heike. *Customer Relationship Management und neue Ansätze*. Wedel: Vorlesung an der Fachhochschule Wedel: [Customer Relationship Management und neue Ansätze], Gehört im Sommersemester 2012.
- Schneider, Prof. Dr. Holger. *E-Commerce Systeme*. Wedel: Vorlesung an der Fachhochschule Wedel: [E-Commerce Systeme], Gehalten im Sommersemester 2013.

Inhalte aus Workshops

- WibiData. *Big Data Workshop*. San Francisco (US): WibiData [Big Data Workshop], 2012.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungskommission vorgelegt und auch nicht veröffentlicht.

Tornesch, 19.06.2013

Ort, Datum

Fabian Hoffmann