

Profilbasierte Methoden für die Empfehlungsgenerierung

Thomas Pernpeintner

12/2004

Zusammenfassung

Ziel dieser Arbeit ist die Untersuchung von Methoden, die zur automatisierten Erstellung von Empfehlungen für Kunden eines Online-Services dienen. Als Daten-Basis für diese Empfehlungen dienen Kunden- und Produktprofile. Diese werden mit den vorgestellten Methoden analysiert, kombiniert und zur automatisierten Generierung von Empfehlungen verarbeitet. Anschließend werden Anwendungsmodelle anhand eines Beispiels aus der freien Wirtschaft vorgestellt.

Inhaltsverzeichnis

1	Einleitung	3
2	Recommendation Systems	3
2.1	Klassifizierung von Recommendation Systems	3
2.2	Ziele und Auswirkungen von Recommendation Systems	4
2.3	Datenschutz	4
3	Methoden für die Empfehlungsgenerierung	5
3.1	Erstellung von Profilen	5
3.2	Collaborative Filtering	7
3.3	Informational Filtering	10
3.4	Kombination der Filterverfahren	10
4	Anwendungsmodelle von Empfehlungsgenerierungen	11
5	Fazit	13

1 Einleitung

Mittlerweile hat sich das World Wide Web als Informations- und Handelssystem in weiten Teilen der Bevölkerung etabliert. Begünstigt wurde dieser Umstand durch den stetigen Anstieg von Internet-Präsenzen, die Besucher mit Angeboten aller Art locken. Mittlerweile haben sich vor allem große Portale durchgesetzt, die Produkte auf breiter Basis anbieten. Ähnlich wie beim Eintreten in großen Kaufhäusern wird der Besucher mit einer breiten Auswahl von Angeboten überflutet. Für den Kunden ist es angesichts der enormen Produktvielfalt oft nicht leicht den Überblick zu bewahren und sein eigentliches Ziel zu finden. Da auch der Inhaber des Online-Shops bzw. Kaufhauses von dem Orientierungsvermögen des Kunden profitiert, sind individuelle Wegweiser für beide Seiten durchaus hilfreich. Im Fall des realen Geschäfts gestaltet sich dieses Vorhaben eher schwierig, da das Vorhaben für jeden Kunden ein Schild mit beispielsweise der Aufschrift „2. Stock, zweite Reihe rechts finden genau Sie Ihr gesuchtes Produkt“ das Erscheinungsbild des Kaufhauses und das Handlungsvermögen der Kunden empfindlich stören würde. Einfacher gestaltet sich hier das Vorhaben für den Inhaber eines Online-Shops, da dieser Shop erst beim Betreten „gebaut“ wird. Weiß der Betreiber ungefähr über die Interessen und Produktwünsche des Kunden Bescheid, so können diese für die Gestaltung des Shops berücksichtigt werden. Im Idealfall entsteht pro Kunde ein Geschäft, das individuell auf ihn zugeschnitten ist, wie auch Jeff Bezos von Amazon.com¹ schon ehrgeizig formulierte:

„If I have 3 million of customers on the Web, I should have 3 million stores in the Web.“

Um an die zur individuellen Gestaltung der Internet-Präsenz benötigten Informationen über die Interessen des Kunden zugelenken dienen Kundenprofile, die durch Beobachtung oder freiwillige Meinungsäußerungen zustande kommen. Im folgenden werden Methoden zur automatisierten Empfehlungsgenerierung auf Basis der Auswertung von Kunden- und Produktprofilen beschrieben, Anwendungsmodelle aufgezeigt und auf wirtschaftlichen Nutzen hingewiesen.

2 Recommendation Systems

2.1 Klassifizierung von Recommendation Systems

Nach Resnick geben in *“einem typischen Recommendation System Personen Empfehlungen ab, die das System speichert und in geeigneter Weise Empfängern präsentiert”* [2]. Diese Definition schliesst drei Klassen von Systemen ein [1]. *“Suggestion Systems”* liefern eine Empfehlungsliste von Produkten die zum Kauf anregen sollen. *“Estimation Systems”* liefern Voraussagen von Benutzerpräferenzen bezüglich spezifischer Produkte, d.h. es wird die Interessen werden spekulativ geschätzt. Diese Art von Systemen dienen vor allem zur individuellen Gestaltung der Website, insbesondere einer auf den Benutzer angepasste Navigation. *“Comment Systems”* verschaffen Zugang zu Empfehlungen, die die Mitglieder in Form eines Kommentars oder eines Bewertungsformular in einer Online Community abgegeben haben. In der vorliegenden Arbeit wird der Begriff *“Recommendation System”* als ein System bezeichnet, das Empfehlungen auf der Datenbasis von Meinungs- und Bewertungsprofilen generiert. Das Recommendation Systems soll seinem Benutzer, im folgenden als **Zielperson** bezeichnet, bei der Auswahl von Produkten unterstützen.

¹Amazon Homepage <http://www.amazon.com>

Dies schliesst eine möglichst individuelle Darstellung der Internetpräsenz sowie eine auf die Interessen der Zielperson ausgerichteten Navigation ein.

2.2 Ziele und Auswirkungen von Recommendation Systems

Ziel der Empfehlungsgenerierung ist es die Effektivität und Effizienz eines E-Commerce Systeme zu steigern. Dies geschieht im Wesentlichen auf drei Arten. Besucher einer Website rufen oft nur Informationen ab ohne dabei Produkte zu bestellen und somit Absatz zu erzeugen. Für den Betreiber bedeutet dies erhöhte Bereitstellungskosten bei niedriger Effizienz. Das Recommendation System hat nun die Aufgabe die Effizienz zu steigern in dem es Interessenten zu Kunden umwandelt. Ein weiteres Ziel ist die Steigerung des sogenannten *Cross-Sellings*². Dabei wird die bestehende Kundenverbindung zum Verkauf weiterer ergänzender Produkte oder Dienstleistungen genutzt, die den Interessen des Kunden möglichst nahe kommen. Die Aufgabe des Recommendation Systems ist hierbei diese Produkte oder Dienstleistungen zu finden. Besteht bereits eine Kundenverbindung hilft die Empfehlungsgenerierung die Kundenloyalität zu steigern. Im Hinblick darauf, dass im weltweiten Netz die Konkurrenz nur ein Mausklick entfernt ist, gehört die Bindung des Kunden zum Online-Portal zur essentiellen Unternehmensstrategie [16]. Je mehr der Kunde mit dem Recommendation System interagiert, d.h. umso umfangreicher sein Benutzerprofil ist, umso höher ist die Bindung zu diesem Kunden. Dies liegt vor allem an dem Umstand, dass selbst wenn ein Mitbewerber ein identisches Produkt- und Leistungsangebot bereit stellt der Kunde viel Zeit und Energie damit verbringen müsste um diesem System seine Präferenzen „anzulernen“ [17].

2.3 Datenschutz

Zur Gewinnung eines Kundenprofils sind die Erhebung, Speicherung und Verarbeitung personenbezogener Daten erforderlich. Somit untersteht der Betreiber datenschutzrechtlichen Rahmenbedingungen. Die Datenschutzgesetze sollen die Persönlichkeit und die Grundrechte von Personen schützen, insbesondere religiöse, weltanschauliche, politische oder gewerkschaftliche Ansichten, Gesundheit und Intimsphäre. Auskunft über die rechtliche Lage in Deutschland bietet das Bundesministerium für Arbeit und Wirtschaft in seinem „*Wegweiser zu den Informations- und Kommunikationsdienste Gesetzen*“ [9]. Im Teledienstschutzgesetz (TDDSG) der Bundesrepublik Deutschland werden die Pflichten eines Diensteanbieters genau festgelegt. So hat er „*den Nutzer zu Beginn des Nutzungsvorgangs über Art, Umfang und Zwecke der Erhebung, Verarbeitung und Nutzung personenbezogener Daten sowie über die Verarbeitung seiner Daten (...) zu unterrichten, sofern eine solche Unterrichtung nicht bereits erfolgt ist. Bei automatisierten Verfahren, die eine spätere Identifizierung des Nutzers ermöglichen und eine Erhebung, Verarbeitung oder Nutzung personenbezogener Daten vorbereiten, ist der Nutzer zu Beginn des Verfahrens zu unterrichten. Der Inhalt der Unterrichtung muss für den Nutzer jederzeit abrufbar sein.*“ [10]. Für die automatisierte Empfehlungsgenerierung bedeutet dies, dass die betroffenen Personen zur Speicherung und Verwertung ihrer Daten ihr Einverständnis erklären. Dies kann in Form einer Einverständniserklärung der allgemeinen Geschäftsbedingungen (AGB), in denen die Verwendung persönlicher Daten erklärt wird, online per Formular geschehen. Somit liegt eine wirksame Einwilligung des Nutzers zur Erstellung eines Persönlichkeitsprofils vor. Soll ein Online-Shop jedoch bereits auf der Startseite Empfehlungen präsentieren, kommt nur die anonymisierte Erhebung von Kundendaten in Frage, insofern gesetzestreu gehandelt werden will. Das TDDSG erlaubt dem Diensteanbieter „*für die Zwecke*

² *Cross-Selling* ist eine Marketing-Bezeichnung einen potentiellen Kunden anzuregen von in Verbindung stehende Produkte oder Dienstleistungen zu kaufen.[15]

der Werbung, Marktforschung oder zu bedarfsgerechten Gestaltung der Teledienste Nutzungsprofile bei der Verwendung von Pseudonymen erstellen, sofern der Nutzer dem nicht widerspricht“.

3 Methoden für die Empfehlungsgenerierung

Der erste Schritt zur automatisierten Empfehlungsgenerierung ist die Auswahl von relevanten Einflussgrößen, die aus verschiedenen Quellen stammen können. Zum einen liefert die Beobachtung der Zielperson bezüglich seines Verhaltens, wie beispielsweise der Häufigkeit von abgerufenen Seiten, Suchbegriffen oder Navigation, relevante Informationen. Zum anderen werden Bewertungseinrichtungen ausgelesen und die Analyse von Beiträgen in Community-Foren als Meßgröße verwendet. Aus diesen Daten werden Produkt- und Kundenprofile generiert die wiederum als Datenbasis für Filterverfahren dienen. Diese Filtermethoden generieren dann die Empfehlung.



PRINZIPIELLER AUFBAU EINES RECOMMENDATION SYSTEM

Heuristik für die Methoden der Empfehlungsgenerierung ist dabei die Annahme, dass Benutzer, die einander ähnlich sind, auch ähnliche Präferenzen haben. Des weiteren wird davon ausgegangen, dass die Benutzer bei der expliziten Meinungsäußerung wahrheitsgetreu handeln.

3.1 Erstellung von Profilen

Jedes Recommendation System benötigt für die Generierung einer Empfehlung als Eingabe eine Reihe von Informationen über die Produkte, der zu beratenden Person sowie über die Gesamtmenge aller Benutzer. Trägt man die Präferenzen, typischen Kennzeichen und Merkmale eines Benutzers zusammen, so erhält man ein individuelles Profil dieses Benutzers. Da für jeden Benutzer die gleiche Vorgehensweise angewendet wird, können die verschiedenen Profile untereinander anhand ihrer Merkmale verglichen werden. Das Profil der Zielperson, ist vom Recommendation System in Echtzeit, dass heisst während der Benutzer mit dem System kommuniziert, zu erstellen. Falls für den Benutzer bereits ein Profil angelegt wurde, wird dieses erweitert oder korrigiert. Das erzeugte Profil wird zu einem der Eingabeparameter, gleichzeitig wird es abgespeichert. Die sich so anhäufenden Benutzerprofile werden, wie später beschrieben, für die Empfehlungsgenerierung als Vergleichsprofile benötigt. Um dabei Redundanz zu vermeiden darf während des Vergleichsprozesses das Profil der Zielperson nicht in der Menge der Vergleichsprofile sein. Neben den Profilen der Benutzer benötigt ein Recommendation System die Kategorisierung der angebotenen Produkte. Dabei wird aus den gemeinsamen Merkmalen der Produkte das Profil erstellt. Beispielsweise bilden in Online-Shop für CDs Titel, Albumname, Erscheinungsjahr, Name des Künstlers oder Interpreten, Genre und Preis ein mögliches Produktprofil. Auch hier garantieren die gemeinsamen Merkmale des Profils die Vergleichbarkeit der Produkte untereinander.

Künstler/Interpret	Album	Genre	Jahr	Preis
Norah Jones	Come away with me	Pop / Jazz	2002	13.49 EUR
Diana Krall	The Look of love	Jazz	2001	13.49 EUR
Ray Charles	The very best of Ray Charles	Jazz	2000	7.99 EUR
Madonna	Like a prayer	Pop	1990	10.99 EUR
Thelonious Monk	Monk's Dream	Jazz / Bepop	1963	10.99 EUR
Britney Spears	My Prerogative	Dance / Pop	2004	13.99 EUR

TABELLE 1: BEISPIEL FÜR DIE DATENBASIS EINES PRODUKTPROFILS

Bei der Erstellung eines Profils ist prinzipiell zwischen impliziten und expliziten Verfahren zu unterscheiden. **Implizite Verfahren** liefern Daten durch Beobachtungen des Benutzers. Beobachter ist im Regelfall das System, welches den Dienst anbietet. Im Falle eines Online-Shops ist dies der Server auf dem der Shop implementiert ist. Bei jeder Anfrage an den Server protokolliert dieser die Aktion in einem Logfile. Somit liegen über jeden Benutzer Informationen vor wann er von welchem Ort welche Seite besucht hat, wie lange er sich darauf aufgehalten und über welchen Weg er diese Seite erreicht hat. Aus diesem Informationspaket lassen sich einige Erkenntnisse schliessen. Da ein Produkt oft einer Seite zugeordnet ist lässt somit feststellen für welche Produkte der Benutzer Informationen abgerufen hat. Produkte sind meist themenspezifischen Kategorien zugeteilt, so dass sich auch die Informationsabrufe der angebotenen Themen quantitativ erfassen lassen.

Benutzer-ID	Timestamp	File	Referrer
?	12.10.2004, 13:15:12	index.html	http://www.google.tld?s=themaX
?	12.10.2004, 13:15:14	login.html	index.html
Bücherkäufer	12.10.2004, 13:17:11	uebersicht.html	login.html
Bücherkäufer	12.10.2004, 13:18:12	fachbuecher/Buch_themaX.html	uebersicht.html
Bücherkäufer	12.10.2004, 13:19:11	uebersicht.html	fachbuecher/Buch_themaX.html
Bücherkäufer	12.10.2004, 13:19:34	krimi/thema_XRoman.html	uebersicht.html
Bücherkäufer	12.10.2004, 13:20:31	uebersicht.html	krimi/thema_XRoman.html
Bücherkäufer	12.10.2004, 13:20:54	fachbuecher/Buch_themaX.html	uebersicht.html
Bücherkäufer	12.10.2004, 13:21:07	bestellung.html	fachbuecher/Buch_themaX.html
Bücherkäufer	12.10.2004, 13:22:31	http://www.shop.tld/logout.html	bestellung.html

TABELLE 2: BEISPIEL FÜR EINEN AUFBEREITETEN LOGFILE (AUSSCHNITT)

Die Tabelle 2 zeigt ein exemplarisches Ausschnitt eines aufbereiteten Logfiles³. Ein Benutzer kommt über eine Suchmaschine durch einen bestimmten Suchbegriff zu dem Onlineshop. Dort zeigt ihm eine Übersichtsseite relevante Produkte, die zu dem Suchbegriff passen. Er ruft eines dieser Produkte in einer Kategorie ab, geht zurück zur Übersicht und ruft ein zweites Produkt einer anderen Kategorie ab. Schliesslich entscheidet er sich scheinbar für das Produkt der ersten Kategorie, das er letztendlich auch bestellt und anschliessend den Online-Shop verlässt. Aus diesem Szenario lässt sich schliessen, dass das zweite Produkt nicht von grösserem Interesse ist, wohl aber das erste Produkt. Diese Informationen fliessen direkt in die Profilbildung ein. Im Falle des Beispiels aus Tabelle 2 würde die Kategorie „Fachbücher“ bezüglich des Suchbegriffs „themaX“ höher gewichtet und somit für den Kunden „Bücherkäufer“ als relevanter eingestuft als die Kategorie „Krimi“. Ein weiteres profilbildendes Element ist der Abschluss eines Kaufvertrags. Entscheidet sich der Benutzer für den Kauf eines Produktes lässt dies allerdings nur auf das Interesse des Käufers am Produkt interpretieren. Eine Bewertung des Produkts durch den Benutzer ist das implizite Verfahren nicht in der Lage zu leisten. Aufgabe ist es viel mehr das Benutzerprofil um Interessensgebiete und um Protokolle bereits abgerufene Produktinformationen zu erweitern. Dem beobachtendem Wesen nach bleibt dem Benutzer ein implizites Verfahren stets gänzlich verborgen.

Im Gegensatz dazu fordern **explizite Verfahren** den Benutzer auf Informationen über sich preiszugeben. Anhand von Bewertungs-Skalen für Produkte können Benutzer ihre Präferenzen angeben. Diese Skalen können sowohl numerischen (z.B. Bewertung nach Schulsystem) als auch binären Charakter („Gefällt Ihnen das Produkt?“) haben. Um möglichst Bewertungen für die gesamte Produktpalette des Online-Shops zu erhalten, bietet sich an die zu bewertenden Produkte und die bewertenden Benutzer zufällig auszuwählen. Dieses Szenario hat dann den Charakter einer Marktumfrage, bei der das Tupel Proband / Bewertungsobjekte unabhängig und somit repräsentativ ist. Motiviert werden kann der Benutzer sich die Zeit zu nehmen um seine Vorlieben und Abneigungen preiszugeben durch Hinweise auf einfachere Navigation und effizientere Bedienung sowie durch neue inhaltliche Anregungen. Insbesondere hierbei ist zu beachten, dass die Angaben der Benutzer aus verschiedenen Motivgründen nicht unbedingt ihrer tatsächlichen Meinung entsprechen. Breitere Akzeptanz findet das aktive Editieren des Profils durch den Benutzer selbst. Die Selbstauskunft über Präferenzen und Abneigung von Produkten oder Produkt-Kategorien bilden den Kernpunkt eines Benutzerprofils, da davon auszugehen ist, dass die explizite Angabe am ehesten die Meinung des Benutzer abbildet. Gerade personalisierte Online-Shops beruhen auf die Profilmomente der Selbstauskunft

³Mehr Informationen und Methoden zum Aufbereiten von Logfiles sind u.a. in [11] zu finden.

kunft. So können die Benutzer von CDNOW⁴ angeben welche Musikalben sie bereits besitzen und wie sie diese bewerten. Hauptaufgabe des expliziten Verfahrens ist die Erweiterung des Profils um Bewertungen und Meinungen des Benutzers über Produkte.

Durch Zusammentragen der Daten aus impliziten und expliziten Verfahren entstehen einheitliche und umfangreiche Profile der Benutzer und der Produkte, welche nun als Datenbasis für die Empfehlungsgenerierung fungieren.

3.2 Collaborative Filtering

Die Empfehlungsgenerierung durch Collaborative Filtering basiert auf Evaluierung von Produkten durch Personen. Die Benutzer eines Online-Auftritts haben die Möglichkeit die inhaltlichen Elementen, bei einem Online-Shop sind das typischerweise die Produkte, Bewertungen abzugeben, die das Recommendation System speichert und zur Generierung von Empfehlungen für andere Benutzer verwendet [6]. In der einfachsten Form werden die einzelnen Evaluierungen generalisiert und zu Empfehlungen zusammengefasst. Konkret heisst das, dass für jedes Produkt die positive oder negative Bewertungen quantitativ erfasst werden und anschliessend eine, nach der Anzahl der Bewertungen geordnete, Rangliste generiert wird. Diese Art von Empfehlung ist jedoch nicht auf das persönliche Profil des Benutzers zugeschnitten. Vielmehr findet es Einsatz falls noch kein ausreichendes Profil für die Zielperson gebildet werden konnte und somit zu wenige Daten für Vergleichsalgorithmen vorliegen, was insbesondere bei Erstbenutzern der Fall ist. Fortgeschrittene Systeme suchen für jeden Benutzer „artverwandte Gruppen“ bestehend aus Personen, deren Meinung stark zu dem des Benutzers korrelieren. Um Ähnlichkeiten zwischen Personen bezüglich ihrer Präferenzen festzustellen, wurden für jeden Benutzer Daten gesammelt und in ein Profil projiziert. Anhand dieses Profils soll nun ermittelt werden welche Benutzer dem Profil am nächsten sind. Um diesen Prozess der „Nachbarschaftssuche“ zu verdeutlichen, wird die Vorgehensweise im folgenden exemplarisch demonstriert. Ein allgemeines Verfahren zur Findung von Produkten, die für die Zielperson von potentiell Interesse sind, und das Vorraussagen trifft wie die Zielperson wahrscheinlich ein Produkt bewertet beschreibt [3]. In Tabelle 3 sind fünf Personen gelistet, die jeweils Bewertungen für fünf Restaurants abgegeben haben. Die Spalten repräsentieren das Profil des Benutzers. Ein „+“ (+1.0) bedeutet eine positive, ein „-“ (-1.0) eine negative Bewertung, wurde noch keine Bewertung abgegeben ist in die entsprechende Zelle in „?“ eingetragen.

	Anton	Berta	Christian	Dominik	Erna
Zur Post	-	+	+	+	-
Antonio	+	+	+	+	+
Chan Thai	+	-	+	-	+
Mykonos	-	+	-	+	-
Mai Ling	+	-	+	-	?

TABELLE 3: BEWERTUNG VON FÜNF RESTAURANTS DURCH FÜNF PERSONEN

Um eine Vorraussage zu treffen wie Erna das Restaurant Mai Ling bewerten würde, werden ähnliche Bewertungsmuster zu Ernas Profil gesucht. In diesem Beispiel haben Anton und Erna den gleichen Geschmack, was zu der Vorraussage führt, Erna bewertet Mai Ling mit einem „+“, da Anton dies genauso beurteilt. Ziel ist es nun einem allgemeingültigen, deterministischen Algorithmus zu formulieren der

⁴CDNOW Homepage <http://www.cdnw.com>

anhand der Benutzerprofile eine Vorraussage trifft, wie eine Person ein noch nicht bewertetes Produkt oder Kategorie (in diesem Beispiel die Restaurants) beurteilt. Zunächst wird hierfür der Grad an Profil-Übereinstimmungen zwischen zwei Personen bestimmt. Dieser Grad an Übereinstimmungen wird als Gewicht für die Relevanz des Vergleichsprofils interpretiert, d.h. der Grad beschreibt den Anteil, zu dem das Vergleichsprofil in die Vorraussage einfließt. Um diesen Grad zu bestimmen wird beim Collaborative Filtering der *Pearson nearest neighbour* Algorithmus verwendet. Der Ausdruck r ist ein Maß für den Grad an Übereinstimmungen von den zwei Personen x und y , da r die „inhaltliche Entfernung“ zweier Personen ermittelt. $R_{i,j}$ ist dabei der Bewertung des Produkt j durch die Person i .

$$r(x, y) = \frac{\sum_{d \in \text{Produkte}} (R_{x,d} - \overline{R_x}) \cdot (R_{y,d} - \overline{R_y})}{\sqrt{\sum_{d \in \text{Produkte}} (R_{x,d} - \overline{R_x})^2 \cdot \sum_{d \in \text{Produkte}} (R_{y,d} - \overline{R_y})^2}}$$

WOBEI $\overline{R_x}$ DER MITTELWERT ALLER BEURTEILUNG VON PERSON X IST.

In diesem Beispiel ist der Grad an Übereinstimmungen zwischen Erna und Anton 1.0, zwischen Erna und Berta -0.577, zwischen Erna und Christian 0.577 und zwischen Erna und Dominik -0.577. Folglich ist das Durchschnittsgewicht des Produktes einer Benutzerbewertung für Mai Ling und der Grad an Übereinstimmungen zwischen Erna und diesem Benutzer 0.682. Ein Collaborative Filtering Algorithmus folgert auf Basis der Bewertungen der anderen Personen, dass Erna das Restaurant Mai Ling positiv bewerten würde. Unterstützt wird diese Einschätzung durch die Tatsache, dass Erna und Dominik einen gegensätzlichen Geschmack haben und Dominik Mai Ling negativ bewertet hat. Ein Recommendation System würde somit Erna empfehlen, das Restaurant Mai Ling zu besuchen.

Collaborative Filtering kann neben der Evaluierung von Personen-Korrelationen auch eingesetzt werden Beziehungen der bewertenden Produkte zu erkennen. Aus Tabelle 3 kann man entnehmen, dass die beiden Restaurants Chan Thai und Mai Ling von jeder Person die gleiche Beurteilung bekamen. Das veranlasst die Annahme, dass diese beiden Restaurants im Auge der Bewerter ähnlich sind. Hier setzt die Empfehlungsgenerierung an: Erna wird Mai Ling empfohlen, wenn sie Chan Thai positiv bewertet hat. Auch dieses Vorgehen lässt sich generalisieren, in dem die Korrelationen der Produkte mit der Pearson'schen nearest neighbour-Kennzahl $r(x, y)$ bestimmt werden. Eine Empfehlung wird dann auf der Basis der gewichteten Durchschnittsbewertungen $\overline{R_x}$ der anderen Produkte erstellt.

Die Effektivität des Collaborative Filtering, also die Qualität der Empfehlung, wird jedoch durch einige Faktoren und Hindernisse vermindert oder sogar gänzlich zur Nichte gemacht. Unter dem *First Rate Problem* [8] versteht man das Unvermögen des Systems eine qualitative Empfehlung für einen neuen oder nicht identifizierbaren Benutzer abzugeben. Grund dafür ist, dass für einen neuen Benutzer noch kein Profil vorliegt, den der Algorithmus für die Berechnung des *nearest neighbour* aber braucht [3]. Analog sinkt die Qualität der Empfehlung mit sinkender Anzahl von Profilen und abgegebenen Bewertungen der Zielperson. Ein umfangreiches Profil der Zielperson und eine hohe Anzahl von Vergleichs-Profilen lässt auf eine gute Qualität der Empfehlung schließen. Allerdings besteht dabei die Gefahr der *Scalability* [13]. Sind Millionen von Benutzerprofilen und Millionen von Produkten im System gespeichert besteht die Gefahr, dass ein typisches webbasiertes Recommendation System, das die Empfehlung für jeden Benutzer und in Echtzeit generieren

muss, durch die enorme Datenmenge und die damit verbundene Rechenlast zusammenbricht. Einen Ausweg bietet effizientere Algorithmen wie das *Clustering*, das im Rahmen dieser Arbeit nicht weiter beleuchtet wird. Es sei an dieser Stelle auf weiterführende Literatur wie beispielsweise [12] verwiesen. Ähnlich dem *First Rate Problem* ist die Voraussetzung für eine Empfehlungsgenerierung, dass für die Produkte Bewertungen vorliegen. Besonders kritisch ist dies bei Online-Systemen, denen ständig neue Produkte hinzugefügt werden, wie es beispielsweise bei einer Online-Zeitung oder Versteigerungssystemen wie ebay⁵ der Fall ist. Diese neuen Produkte können per *Collaborative Filtering* nur dann empfohlen werden, wenn bereits Bewertungen seitens der Benutzer abgegeben worden sind. Ein Ausweg bietet hier eine andere Vorgehensweise zur profilbasierten Empfehlungsgenerierung: das *Informational Filtering*.

3.3 Informational Filtering

Zu den ältesten Vertreter von Recommendation Systems zählt das Informational Filtering. Ziel ist die Vermeidung von Reizüberflutung und eine daraus resultierende hohe Kundenfluktation verursacht durch die Vielzahl von angezeigten Informationen auf der Internetpräsenz. Dies wird realisiert indem gezielt nur die Informationen eingeblendet werden, die den Interessen des Betrachters oder aber auch den Interessen des Anbieters entsprechen. Um Entscheiden zu können welche Informationen relevant sind bietet das Informational Filtering verschiedene Ansätze an.

Als zentrale Datenquelle dient dem Informational Filtering die sogenannte *item to item* Korrelation. Sie dient dazu Produkte (*items*) zu finden die in Assoziation zu Produkten steht, für die der Kunde bereits Interesse gezeigt hat. Assoziationen werden auf Grundlage von Verkaufsdaten, Präferenzen der Allgemeinheit und weiteren Merkmalen des Produktprofils gebildet. In der einfachsten Form wird die *item to item* Korrelation benutzt um für ein einzelnes Produkt verwandte Produkte, d.h. Produkte mit einer bestehenden Assoziation zum dem einzelnen Produkt, zu finden. Komplexere Systeme generieren zu einer Menge von Produkten, wie sie beispielsweise in einem Einkaufskorb vorliegen, relevante Empfehlungen.

Das sogenannte „**content based informational filtering**“ ist eines der am häufigsten implementierten Techniken. Sie untersucht Texte auf ihren buchstäblichen Inhalt und vergleicht diesen mit dem Interessenprofil der Zielperson [18]. Die *item to item* Korrelation ist die zentrale Datenquelle für diese Technik. Annahme ist hierbei, dass Kunden mit ähnlichen Interesse sich für ähnliche Produkte interessieren. Dies ist allerdings oft schwer einschätzbar, da sich die Frage stellt wie definiert sich „ähnlich“. Beispielsweise sind ähnliche Filme an Kriterien wie gleiche Schauspieler, gleiches Genre, gleicher Regisseur erkennbar. Ziel ist es also Produkte durch wesentliche Kriterien zu beschreiben, wobei die Kriterien mit denen anderer Produkte vergleichbar sein müssen. Somit ist eine Gruppenbildung von ähnlichen Produkten möglich. Die wesentliche Annahme der *item to item* Korrelation kann durch folgende Aussage beschrieben werden: „Wen jemandem ein Produkt aus einer Gruppe gefällt dann interessiert er sich wahrscheinlich auch für das ein anderes Produkt dieser Gruppe“. Das *content based informational filtering* zeigt der Zielperson nur Informationen, die dem beobachtetem Interesse „ähnlicher“ Personen entsprechen.

Eine weitere Technik des Informational Filtering ist das „**demographic based filtering**“, unter der man Empfehlungen von Produkten versteht, die von demographisch ähnlichen Personen besichtigt oder durch Bewertung empfohlen wurden. Die Vergleichs-Kriterien im Benutzerprofil sind demographische Daten wie Geschlecht, Alter, Region, Beruf, etc. Dazu müssen diese Daten bekannt sein, was bei vielen

⁵Ebay Homepage <http://www.ebay.de>

Internetplattformen nicht der Fall ist. Grund dafür ist oft mangelnde Bereitschaft von Besuchern diese Daten freiwillig und vor allem korrekt anzugeben.

Die „**economic informational filtering**“ - Technik basiert auf verschiedenen Arten von Kosten-Nutzen-Kalkülen, in denen Informationen als ein Gut eines mikroökonomischen Marktmodells angesehen wird [19]. Mit Hilfe expliziter und impliziter Preismechanismen wird die Informationsmenge gefunden, in dem sich die Kosten der Information und der erwartete Nutzen entsprechen. Die in die Kalkulation einbezogene Kosten müssen dabei nicht den monetären Kosten entsprechen. Dieses Thema wird vertiefend in [20] behandelt.

3.4 Kombination der Filterverfahren

Die beiden Methoden Informationale Filtering und Collaborative Filtering können jeweils auf sich allein gestellt nicht in jeder Situation eine relevante Empfehlung generieren. Da die beiden Methoden jedoch auf zwei sehr unterschiedlichen Ansätzen beruhen und einen eher entgegengesetzten Charakter haben bietet sich als Ausweg ein Hybridsystem an, in die die beiden Verfahren integriert werden. So gleichen die Stärken der einen die Schwächen der anderen Methode aus [1]. Allerdings ist es auf Grund der grossen Datenmengen und der verschiedenen Verarbeitungsmethoden von Teilmengen dieser Daten sinnvoll, diese nach ihren Quellen getrennt behandeln. Ziel ist es die Ergebnisse der beiden Methoden zusammenzuführen und als einheitliches Empfehlungsinstrument zu präsentieren wobei auf mögliche Redundanzen, Widersprüche und Wiederholungen zu achten ist. Das Recommendation System Krakatoa [21] löst dies in dem die Benutzer die Verhältnismässigkeit des Einflusses der Ergebnisse von Informationale und Collaborative Filtering per virtuellem Schieber online selbst bestimmen können. Ein ähnlicher Ansatz ist es die einzelnen Ergebnisse der Filterverfahren zu gewichten und sie anhand dieser Gewichtungen zu kombinieren. Da die Ergebnisse der Filterverfahren jedoch sehr unterschiedlichen Charakter haben ist die Vergleichbarkeit der Gewichtungen eher eingeschränkt. Aktuelle Recommendation Systems, wie sie beispielsweise derzeit bei Amazon⁶, CDNOW⁷ oder Drugstore⁸ eingesetzt werden, kategorisieren ihre Empfehlungen nach der Art und somit auch nach dem Verfahren, das sie generiert hat. So sind die Sparten wie „Kunden die dieses Produkt gekauft haben, haben auch diese Produkte gekauft“, „Meist verkauftes Produkt“ und „Beliebteste Kategorie“ auf Informationale Filtering zurückzuführen. Empfehlungen die durch die Methoden des Collaborative Filtering entstanden sind typischer Weise Benutzer-Bewertungen durch Punkte, Einblendungen von relevanten Kommentaren (sogenannte „Reviews“) und Links zu Produkten und Kategorien die bereits von Personen ähnlichen Profils als interessant eingestuft worden sind. Die explizite Kennzeichnung der Quelle der Empfehlungen schafft auch mehr Transparenz und somit auch höhere Akzeptanz bei der Zielperson.

4 Anwendungsmodelle von Empfehlungsgenerierungen

Der *State of the Art* von Recommendation Systems lässt sich anhand der bereits genannten Online-Shop wie Amazon, CDNOW oder Drugstore erkennen. Sie alle ähneln sich in ihrer Umsetzung und in der Präsentation der Empfehlungen. Daher wird im folgenden der Online-Shop für Musikalben CDNOW als Repräsentant näher untersucht. Die Produktpalette und die Struktur des Internetportals eignet sich besonders gut für den Einsatz eines Recommendation System. Die Website passt

⁶Amazon Homepage <http://www.amazon.com>

⁷CDNOW Homepage <http://www.cdnw.com>

⁸Drugstore Homepage <http://www.drugstore.com>

sich dem individuellen Geschmack des Kunden an und empfiehlt Alben, die er noch nicht besitzt aber die durch Interpretation des Kundenprofils von Interesse sein dürften. Dazu haben die Benutzer die Möglichkeit Alben anzugeben, die sie bereits besitzen und welche Interpreteten sie favorisieren. Im Falle einer Bestellung wird die sogenannte „own it“ - Liste automatisch erweitert. Obwohl die „own it“ - Liste eher dispositiver als positiver Indikator für das jeweilige Produkt interpretiert wird, kann die Bewertung jederzeit korregiert werden in „own it and like it“ oder „own it and dislike it“.

Der Kunde ruft eine Seite zu einem Album ab. Neben Informationen zu diesem Album und einer Bestellmöglichkeit werden dem Benutzer acht verschiedene Kategorien von Vorschlägen und Empfehlungen präsentiert. Zunächst wird der Zielperson ein „**Better Together**“ Angebot gemacht: Bestellt er neben dem angezeigten Album ein weiteres, bekommt er das Doppelpack zu einem attraktiven Tagespreis, in etwa zwei drittel der Summe der Einzelpreise. Das zusätzliche Album ermittelt das Recommendation System. Die verwendete Technik entspricht dem „Economic Informational Filtering“.

Better Together

Buy this album with [The Very Best of Marvin Gaye](#) ~ Marvin Gaye today!



Total List Price: \$35.96
Buy Together Today: \$30.98
[Buy both now!](#)

In der vom Recommendation-Pionier Amazon⁹ bereits bekannten Rubrik „**Customers who bought this title also bought:**“ sind Alben aufgelistet, die von der Mehrzahl der Kunden gekauft wurden, die auch das Album der angezeigten Seite gekauft haben.

Customers who bought this title also bought:

- [Al Green - Greatest Hits](#) ~ Al Green ([Rate it](#))
- [The Very Best of Otis Redding, Vol. 1](#) ~ Otis Redding ([Rate it](#))
- [Stevie Wonder - The Definitive Collection](#) ~ Stevie Wonder ([Rate it](#))
- [Wilson Pickett's Greatest Hits \[1985\]](#) ~ Wilson Pickett ([Rate it](#))
- [Sly & the Family Stone - Greatest Hits \[Epic\]](#) ~ Sly & the Family Stone ([Rate it](#))
- [Ultimate Collection](#) ~ Temptations ([Rate it](#))
- [Old School Jams: The History Lesson](#) ~ Various Artists ([Rate it](#))
- [Barry White - All-Time Greatest Hits](#) ~ Barry White ([Rate it](#))

► [Explore Similar Items:](#) [20 in Music](#), and [1 in DVD](#)

Ähnlich wie bei der Auflistung von Alben Genre-verwandeter Künstler in der Rubrik „**Customers who bought titles by XY also bought titles by these artists:**“ wird auf Grundlage der „*item to item*“ - Korrelation ein „*content based informational filtering*“ umgesetzt.

Customers who bought titles by James Brown also bought titles by these artists:

- [Funkadelic](#)
- [The JB's](#)
- [Stevie Wonder](#)
- [Ray Charles](#)
- [Sly & The Family Stone](#)

► [Explore similar artists](#)

⁹Amazon Homepage <http://www.amazon.com>

Zum einen schlägt es auf dieser Seite zehn weitere Alben vor. Analog existiert ein Verfahren mit Kategorien. Stöbert die Zielperson in den verschiedenen Musik-Genres werden weitere Genres in der Rubrik „**Browse for music**“ vorgeschlagen, die auf Grund der Profildatenbasis als verwandt interpretiert werden.

Browse for [music](#) in:

- [Styles](#) > [R&B](#) > [Funk](#) > [General](#)
- [Styles](#) > [R&B](#) > [Classic R&B](#)
- [Styles](#) > [R&B](#) > [General](#)
- [Styles](#) > [R&B](#) > [Soul](#) > [General](#)

In der Sparte „**Customers interested in XY may also be interested in:** (...)“ werden zu dem betrachtenden Album themenverwandte externe Links gelistet. Diese Links führen beispielsweise zu einem Kartenverkauf für ein Konzert des Künstlers, zu einem T-Shirt und Posterverkauf oder zu einem Anbieter von Videos. CDNOW generiert die Assoziationen allerdings nicht selbst sondern nutzt das AdSense Programm von Google¹⁰.

Customers interested in James Brown may also be interested in:

Sponsored Links ([What's this?](#)) [Feedback](#)

- [See James Brown on LAUNCH](#)
Watch **James Brown** Music Videos Now! Free Photos, News & More at LAUNCH
[launch.yahoo.com](#)

Neben den Empfehlungen auf Grundlage des *Informational Filtering* verwendet CDNOW auch das *Collaborative Filtering* für sein Recommendation System. Für die Datengrundlage geben die Benutzer ihre Meinung zu den verschiedenen Alben und Künstlern ab. Dazu können sie einen mehrzeiligen Bewertungstext über das Produkt in einem „**Review**“ verfassen.

Our Customers' Advice

See what customers recommend in addition to, or instead of, the product on this page.

▶ [Recommend an item!](#)

- **1 person recommended** [Transformed Man](#) in addition to James Brown - 20 All-Time Greatest Hits!
- **1 person recommended** [Insatiable](#) instead of James Brown - 20 All-Time Greatest Hits!

▶ [See more customer buying advice](#)

▶ [Let us know](#) if any of these recommendations are inappropriate.

Da nicht viele Benutzer sich die Zeit nehmen wollen einen ausführlichen Text zu schreiben bietet CDNOW auch die Möglichkeit anhand der „**Rate It**“ - Funktion, die zu jedem Produkt angezeigt wird, eine Bewertung in fünf Stufen abzugeben. Diese Funktion ist für das Recommendation System von doppelten Nutzen. Zum einen liegen subjektive Wertungen des Produkts vor, zum anderen fließt die Bewertung in das Profil des Zielkunden ein. Die *Rate It* Funktion wird aber nicht nur für Produkte angeboten, sondern auch für die abgegebenen *Reviews*. Somit kann das Recommendation System entscheiden ob von Benutzer abgegebene Kommentare von anderen Benutzern als hilfreich eingestuft wurden.

¹⁰Mehr Informationen zu AdSense von Google <https://www.google.de/adsense/>

5 Fazit

Die profilbasierte Empfehlungsgenerierung ist eine mächtige neue Technologie, die insbesondere für Online-Shops eine Wertsteigerung ermöglicht. Sie hilft dem Kunden schnell das Produkt seines Interesses zu finden bzw. inspiriert neue Produkte kennenzulernen. Der entscheidende Vorteil von automatisierten Empfehlungen ist die erhöhte Kundenbindung. Hoher personeller Aufwand für Kundenberatung entfällt. Eine Bereitstellung individueller Empfehlungen auf Internetportalen durch Personal sind meist nicht wirtschaftlich, deshalb sind automatisierte Methoden zur Generierung von individuellen Empfehlungen der einzige Ausweg. Da die Akquisition neuer Kunden erfahrungsgemäß aufwändiger als der Verkauf weiterer Produkte ist, kann hier mit kleinerem Aufwand ein höherer Nutzen erzielt werden, als beim Verkauf des gleichen Produktes an einen Neukunden. Der Grad der Empfehlung ist nach der Häufigkeit der Besuche zu wählen. Es empfiehlt sich ein breiter Empfehlungsgrad, beispielsweise eine grobe Übersicht der beliebtesten Produkte und Kategorien, für neue und seltene Benutzer. Für stark frequentierende Kunden sind sehr detaillierte Empfehlungen einzusetzen. Eine der wesentlichen Aufgaben der Empfehlungsgenerierung ist die Ermittlung von relevanten Produkten und Dienstleistungen für das Cross-Selling. Problematisch bezüglich Datenschutz ist der Umgang mit den Kundendaten. Dem Benutzer muss klar mitgeteilt werden, dass das zur Empfehlungsgenerierung eingesetzte System sich auf Kundenprofile stützt, und er muss die Erhebung, Speicherung und Verarbeitung seiner Daten ausdrücklich bewilligen. Ausserdem ist ein Teil des automatisierten Systems auf wahrheitsgetreue Angaben der Benutzer angewiesen.

Literatur

- [1] Schafer, J.B., Konstan, J., Riedl, J. Electronic Commerce Recommender Applications. *Journal of Data Mining and Knowledge Discovery*, vol. 5 nos. 1/2, pp. 115-152. (<http://www.grouplens.org>)
- [2] Resnick, P., et al. (1994). *GroupLens: An Open Architecture for Collaborative Filtering of Netnews.*, Chapel Hill, NC, ACN Press, CSCW-94 pp. 175-186
- [3] Pazzani, M. (1999). A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence Review*. 13(5-6): 393-408 (<http://www.ics.uci.edu/pazzani/Publications/Publications.html>)
- [4] Zambetta, F., Abbattista, F., Degemmis, M., Fanizzi, N., Licchelli, O., Lops, P., Semeraro, G., Zambetta, F. Learning User Profiles for Content-Based Filtering in e-Commerce, In *Proceedings AI*AI Workshop su Apprendimento Automatico: Metodi e Applicazioni*. Sienna, Italy. 2002.
- [5] Abbattista F., Degemmis M., Licchelli O., Lops P., Semeraro G., Zambetta F. Improving the usability of an e-commerce web site through personalization. In F. Ricci and B. Smyth (Eds.), *Recommendation and Personalization in eCommerce*, *Proceedings of the Workshop on Recommendation and Personalization in Electronic Commerce*, 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems, 20-29, Malaga, Spain, May 28, 2002.
- [6] Forstinger H. E., *Analyse gegenwärtiger Suchdienste und Konzepte für künftige Wissensauffindung, Kapitel 4 Recommendation Systems*, Diplomarbeit von Harald Erich Forstinger, Juli 1999 <http://www.iicm.edu/thesis/hforstinger/Kapitel>
- [7] *Recommendation Engines*, Projektarbeit Lars Diestelhorst, 23. Juli 2001, Betreuung: Prof. Dr. Florian Matthes, Arbeitsbereich Softwaresysteme, Technische Universität Hamburg-Harburg, (<http://www.sts.tu-harburg.de/papers/2001/Dies01.pdf>)
- [8] Good, Nathaniel, Schafer, J. Ben, Konstan, Joseph A., Borchers, Al, Sarwar, Badrul M., Herlocker, Jonathan L., und Riedl, John. *Combining Collaborative Filtering with Personal Agents for Better Recommendations* , 1999.
- [9] Bundesministerium für Wirtschaft und Arbeit: „Wegweiser zu den Informations- und Kommunikationsdienste-Gesetzen“ im Internet unter <http://www.ii-d.de/iukgd/iukd1.html>, Zugriff 10/2004
- [10] Bundesministerium für Wirtschaft und Arbeit: „Teledienstdatenschutzgesetz“ vom 22.Juli 1997, im Internet unter <http://www.iid.de/iukdg/gesetz/tdd-schutzgesetz.pdf>, Zugriff 10/2004
- [11] Pernpeintner T., Droste M., *Website Monitoring, Möglichkeiten der Informationsgewinnung und betrieblichen Auswertung durch Logfileanalyse und web-basiertes Datamining*, 2003
- [12] Sarwar, B. M., Konstan, J. A., Borchers, A., Herlocker, J., Miller, B., Riedl, J. Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering Systems. Aus *Proceedings of CSCW '98*, Seattle, WA, 1998

- [13] Badrul M. Sawar, Georg Karypis, Joseph Konstan, John Riedl, GroupLens Research Group / Army HPC Research Center, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA
- [14] Sawar, B. M., Karypis, G., Konstan, J. A., Riedl, J., Analysis of Recommendation Algorithms for E-Commerce. Aus Proceeding of the ACM EC'00 Conference. Minneapolis, MN., 2000, Seiten 158-167
- [15] SearchCRM.com, cross-sell - a Whatis.com definition, http://searchcrm.techtarget.com/sDefinition/0,sid11_gci942946,00.html, Abgerufen am 26.11.2004
- [16] Reichheld, F., Loyalty-Based Management. Harvard Business School Review, 1993 (2): Seiten 64-73
- [17] Pine, B. J., Peppers, D., Rogers, M., 1995. Do you want to keep your customers forever? Harvard Business School Review, 1995 (2), Seiten 103-114
- [18] Sackmann, S., Social Informational Filtering, Institut für Informatik und Gesellschaft, Universität Freiburg i. Br., 1996/1997
- [19] Ferguson, I. A., Karakoulas, G. J., „SIGMA: Multiagent Learning and Adaption in an Informational Filtering Market“, 1996
- [20] Malone, T. W., „Intelligent Information-Sharing Systems“, Communication of the ACM, 12/1992, Vol 35, No 35, Seite 49f
- [21] Khamba, T., Bharat, K., Albers, M. The Krakatoa Chronicle - An Interactive, Personalized, Newspaper on the Web. Fourth International World Web Web Conference, Seiten 159-170 ff.