

Interaktive Empfehlungsgenerierung mit Hilfe latenter Produktfaktoren

Benedikt Loepp, Tim Hussein, Jürgen Ziegler

Universität Duisburg-Essen

Zusammenfassung

In diesem Beitrag beschreiben wir ein Verfahren zur Generierung interaktiver Empfehlungsdialoge auf Basis latenter Produktfaktoren. Der Ansatz verbindet auf neuartige Weise Methoden zur automatischen Generierung von Empfehlungen mit interaktiven, explorativen Methoden der Produktsuche. Das vorgestellte Verfahren nutzt verborgene Muster in Produktbewertungen („latente Faktoren“) und erzeugt auf dieser Basis visuelle Dialoge, die den Nutzer schrittweise und intuitiv durch einen Explorationsprozess führen. In einer Nutzerstudie konnten wir zeigen, dass ein derartiger interaktiver Empfehlungsprozess hinsichtlich des Aufwandes und der Zufriedenheit mit den erzielten Resultaten eine deutliche Verbesserung gegenüber rein manuellen oder rein automatischen Verfahren bieten kann.

1 Einleitung

Eine zentrale Herausforderung bei heutigen Webangeboten besteht darin, den Nutzer durch die häufig unübersichtlich große Menge an Alternativen zu denjenigen hinzuführen, die seinen Bedürfnissen und Interessen bestmöglich entsprechen. Dies kann sich auf Handelsgüter, Nachrichtenartikel, Medienartefakte oder beliebige andere Objekte beziehen. Um dieser Problematik zu begegnen, wurden *Empfehlungssysteme* (Recommender Systems) mit dem Ziel entwickelt, aus einer Vielzahl von Alternativen automatisiert eine Untermenge zu selektieren und zu präsentieren, die für den Nutzer von besonderem Interesse oder Nutzen sein könnte (Ricci et al., 2010). Dabei sind aus Nutzersicht neben individuellen Präferenzen auch Aspekte wie die Diversität und der Neuigkeitswert für den Nutzer zu berücksichtigen.

Die kontinuierliche Optimierung von Empfehlungsalgorithmen hat zwar zu einer graduell ständig verbesserten Passung der empfohlenen Items mit dem jeweiligen Nutzerprofil geführt, bringt jedoch gleichzeitig eine Reihe von Problemen mit sich, die zunehmend kritisch diskutiert werden (Konstan & Riedl, 2012). Wesentliche Kritikpunkte sind z.B. die verminderte Kontrolle des Nutzers über den Selektionsprozess oder die geringe Transparenz bezüglich der Ursachen für die gegebenen Empfehlungen (Sinha & Swearingen, 2002). Insbesondere wird aber die mit der Zeit ständig zunehmende und sich selbst verstärkende Eingrenzung der Empfehlungen auf die bisherigen Präferenzen des Nutzers diskutiert, die es diesem potenziell erschwert, neue Themenfelder oder Produktarten zu erschließen (Iacobelli, Birnbaum & Hammond, 2010) oder besser auf situative Bedürfnisse zu reagieren (Chi, 2004). Das populäre Schlagwort der „Filter Bubble“ (Pariser, 2011) kennzeichnet dieses Dilemma.

Pu et al. (2012) sehen erhebliches Verbesserungspotenzial in der Entwicklung nutzerfreundlicher und situationsangemessener Benutzungsschnittstellen für Recommender. Die verwendeten Algorithmen werden hingegen als bereits sehr ausgereift erachtet, so dass hier nur moderate Qualitätssteigerungen zu erwarten sind. Zudem konnte in Studien gezeigt werden, dass Nutzer eine aktivere Rolle bei der Empfehlungsgenerierung wünschen (Xiao & Benbasat, 2007). Ein höherer Einfluss kann die Transparenz des Prozesses steigern und somit die Akzeptanz der präsentierten Ergebnisse erhöhen (Xiao & Benbasat, 2007).

Der vorliegende Beitrag stellt ein neuartiges Verfahren vor, mit dem interaktive Empfehlungsdialoge automatisiert auf Basis existierender Kauf- oder Bewertungsdaten erzeugt werden. Dies soll möglich sein, ohne dass der Nutzer selbst bereits Bewertungen abgegeben hat bzw. ein Nutzerprofil von ihm vorliegt. Die von anderen Nutzern abgegebenen Bewertungen werden mit Hilfe algorithmischer Verfahren auf latente Bewertungsmuster untersucht, um die Produkte anhand der gewonnenen latenten Faktoren zu klassifizieren. Unser Ansatz erzeugt daraufhin automatisiert visuelle Dialoge, um die Interessen des Nutzers zum Zweck der Empfehlungsgenerierung inkrementell zu erfassen. Latente Faktorenmodelle werden seit Jahren erfolgreich in vollautomatischen Empfehlungsverfahren eingesetzt, allerdings sind diese bislang nicht durch den Nutzer steuerbar. Mit unserem Ansatz verbinden wir die Vorteile automatisierter Verfahren (präzise Empfehlungen, kognitive Entlastung) mit denen manueller Exploration (hohe Flexibilität, situative Anpassung und gute Steuerbarkeit).

2 Automatische und interaktive Empfehlungssysteme

Systemgesteuerte, algorithmische Empfehlungssysteme (Ricci et al., 2010) können den Suchaufwand des Nutzers und seine kognitive Belastung reduzieren, sind aber vom Nutzer nicht oder nur wenig steuerbar. Auch ist die Empfehlungsgenerierung häufig nicht oder nur in sehr geringem Maße transparent, was die Akzeptanz der Vorschläge oder das Vertrauen in den Anbieter des Systems negativ beeinflussen kann (Xiao & Benbasat, 2007). Nutzergesteuerte Such- und Filterverfahren sind demgegenüber flexibel und ermöglichen ein exploratives und situativ angepasstes Vorgehen. Andererseits erhöhen sie den Such- und Navigationsaufwand und bieten häufig nicht die zum Nutzerziel am besten passenden Filtermöglichkeiten an. Die in der Praxis eingesetzten Methoden sind zudem vorwiegend auf hierarchische Navigation, Suchfelder und facettierte Filterung beschränkt. Sie setzen voraus, dass der Nutzer sich zumindest ansatzweise ein Suchziel gebildet hat.

Ein vielversprechender Ausgangspunkt zur Überwindung der Probleme und Einschränkungen der jeweiligen Ansätze kann darin gesehen werden, interaktive Methoden mit automatisierten Empfehlungsmechanismen zu kombinieren, zu integrieren und dadurch dem Nutzer eine stärkere Kontrolle über den Empfehlungsprozess zu ermöglichen. Dialogbasierte Empfehlungssysteme konfrontieren den Nutzer beispielsweise mit einer Reihe von Fragen zu seinem Suchziel und generieren anschließend passende Empfehlungen (Mahmood & Ricci, 2009). Die präsentierten Fragen basieren jedoch in der Regel auf vorab modelliertem Wissen, so dass derartige Systeme nur bedingt als dynamisch anzusehen sind.

Sogenannte Critique-based Recommender (Chen & Pu, 2012) erhöhen den Grad an Interaktivität, indem sie den Nutzer die Empfehlungen im Anschluss an ihre Erzeugung in Bezug auf bestimmte Features bewerten lassen. So kann er etwa explizit angeben, dass er eher an preisgünstigeren Produkten interessiert ist. Die Features umfassen in der Regel jedoch nur vorab modellierte Dimensionen. Eine Ausnahme stellt hier der *MovieTuner* von Vig et al. (2011) dar, welcher auf einer großen Menge von durch die gesamte Nutzerbasis vergebenen Tags basiert. *MovieTuner* erlaubt es seinen Anwendern, explizit Empfehlungen von z.B. weniger brutalen Filmen anzufordern. Die für den jeweiligen Film relevanten Tags werden dabei vom System automatisch aus der Gesamtmenge aller Tags herausgefiltert und gewichtet. Der Ansatz ist allerdings nur bedingt übertragbar, da eine große Menge von Tags oder textuellen Beschreibungen über die zu empfehlenden Inhalte benötigt wird.

Bisher existieren nur wenige Ansätze, die Empfehlungssysteme mit interaktiven visuellen Darstellungen kombinieren. Gretarsson et al. (2010) präsentieren eine graphbasierte interaktive Visualisierung für ein in Facebook eingebettetes Empfehlungssystem. In der Evaluation kommen die Autoren zu dem Schluss, dass dies die Transparenz des Systems erhöht, den Empfehlungsprozess leichter verständlich macht und somit die Zufriedenheit der Nutzer verbessert. Zu einem ähnlichen Ergebnis kommen Bostandjiev et al. (2012), welche mit *TasteWeights* ein visuelles, interaktives und hybrides Empfehlungssystem vorstellen. Der Nutzer kann hier den Einflussgrad von Informationen aus verschiedenen sozialen Netzen auf den Empfehlungsprozess beeinflussen, was zu einer Verbesserung der wahrgenommenen Empfehlungsgüte führte.

Zusammenfassend ist festzustellen, dass es bislang nur vereinzelte oder eingeschränkte Ansätze zur Integration von interaktiver Exploration mit automatischer Empfehlungsgenerierung gibt und diese hohe Anforderungen an die zugrundeliegende Datenbasis stellen. In allen Fällen werden reichhaltig vorstrukturierte Daten, zusätzliche Beschreibungen oder Tags benötigt, um dem Nutzer eine Steuerung des Empfehlungsprozesses zu ermöglichen. Verfahren, die ohne diese Voraussetzungen auskommen und den Nutzer trotzdem effektiv und effizient unterstützen, existieren nach unserem Kenntnisstand jedoch nicht.

3 Generierung interaktiver Empfehlungsprozesse

Das im Folgenden vorgestellte Verfahren bestimmt zunächst aus Kauf- oder Bewertungsdaten einer großen Menge von Nutzern eine Anzahl niedrigdimensionaler Faktoren und nutzt diese, um eine Folge von interaktiven Auswahlritten mit typischen Produktvertretern für die unterschiedlichen Faktorausprägungen zu generieren. Nach Durchlaufen einer wählbaren Zahl von Auswahlritten werden dem Nutzer Empfehlungen präsentiert, die bestmöglich zu den von ihm getroffenen Auswahlentscheidungen passen. Ohne Beschränkung der Allgemeinheit wird das Verfahren am Beispiel eines Filmportals erklärt, für dessen Filme eine große Zahl von Nutzerbewertungen von 1 („gefällt mir gar nicht“) bis 5 („gefällt mir sehr gut“) abgegeben wurden. Dies entspricht der Situation bei der gängigsten Empfehlungsmethode, dem Collaborative Filtering.

Ausgangspunkt und einzige Vorbedingung für das Verfahren ist das Vorliegen von Kauf- oder Bewertungsdaten in der Form einer (sehr großen) Nutzer \times Item-Matrix (Items seien hier die zur Auswahl stehenden Filme). Aus den vorliegenden Bewertungsdaten wird eine einge-

schränkte Zahl von Dimensionen, sogenannten latenten Faktoren, gewonnen, anhand derer sämtliche Filme in einem von diesen Dimensionen aufgespannten Vektorraum angeordnet und dadurch charakterisiert werden können. Zur Bestimmung der latenten Faktoren werden Methoden der *Matrix Factorization* (Koren et al., 2009) genutzt, wobei spezifische Verfahren wie *Alternating Least Squares* oder *Stochastic Gradient Descent* herangezogen werden. Für die Einzelheiten dieser Verfahren sei an dieser Stelle auf die entsprechende Literatur verwiesen (Koren et al., 2009).

Die Anzahl der latenten Faktoren, mit denen sich eine Objektmenge optimal klassifizieren lässt, kann variieren – 20 bis 100 Faktoren lassen sich hierbei üblicherweise als guter Richtwert ansehen. Die Faktoren können mehr oder weniger offensichtlich sein, z.B. können sie bestimmte Charakteristika der Filme, etwa „Filme mit schwarzem Humor“ oder „Filme mit romantischer Love-Story“ repräsentieren. Da es sich um ein rein Statistik-basiertes Verfahren handelt, ist es nicht möglich, die Faktoren automatisiert mit Bezeichnungen zu versehen. Es kann lediglich für jeden Film gesagt werden, wie hoch oder niedrig seine Ausprägung bezüglich dieses Merkmals ist.

Die zu erzeugenden Empfehlungen werden anschließend anhand der Ähnlichkeit zwischen den aus den gewonnenen Dimensionen bestehenden Item-Vektoren und dem korrespondierenden Vektor des Nutzers, der seine Interessen repräsentiert, bestimmt. Dieser Nutzervektor ist allerdings zunächst unbekannt, da wir keine im Voraus durch den Nutzer abgegebenen Bewertungen voraussetzen. Stattdessen nutzen wir den erstellten Vektorraum als Basis für Benutzerdialoge, um den Nutzer (bzw. seine Interessen) schrittweise *interaktiv* in diesem Raum zu positionieren. Hierfür wird eine Folge von Auswahlentscheidungen generiert, bei denen der Nutzer jeweils zwischen zwei Mengen an Filmen wählt. Anschließend können wir ihm diejenigen Filme empfehlen, die im Objektraum die geringste Distanz zu seiner Position aufweisen. Abbildung 1 veranschaulicht das Vorgehen an einem einfachen Beispiel mit zwei Faktoren (jeweils repräsentiert durch eine Achse des Vektorraumes).

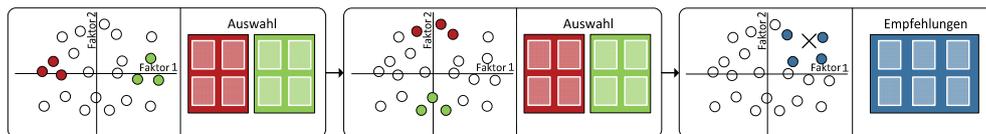


Abbildung 1: Beispielhafter Ablauf des interaktiven Dialog-Prozesses zur Empfehlungsgenerierung bei zwei latenten Faktoren: Für jeden Faktor werden Filme möglichst unterschiedlicher Ausprägung selektiert, die sich in allen anderen Dimensionen sehr ähnlich sind. Die entsprechenden Repräsentanten werden dem Nutzer im Auswahl-dialog präsentiert. Nachdem sich der Nutzer in jedem Schritt für eine Menge entschieden hat, können ihm anhand seiner somit bestimmten Position im Raum Empfehlungen angeboten werden.

In jedem Interaktionsschritt wird implizit die Position des Nutzers gemäß eines Faktors ermittelt, indem ihm zwei kleine Mengen von Filmen präsentiert werden und er nach seiner Präferenz gefragt wird. Während die eine Menge aus typischen Repräsentanten mit geringer Ausprägung eines Merkmals besteht, weisen die Repräsentanten der anderen Menge eine hohe Ausprägung des Merkmals auf. Wählt er die Menge mit den niedrigen Merkmalsausprägungen, erhält der Präferenzvektor bezüglich dieser Dimension einen niedrigen Wert und umgekehrt einen hohen Wert, wenn er die Menge der Repräsentanten mit hoher Merkmalsausprägung wählt. Möchte er für einen Faktor keine Entscheidung treffen, wird diese Dimension nicht weiter berücksichtigt. Diese Vorgehensweise erlaubt eine situativ angepasste Generierung von Empfehlungen und lässt sich als „Recommending by Example“ umschreiben.

3.1 Auswahl und Reihenfolge der Interaktionsschritte

Bei n Faktoren wären theoretisch n Auswahlentscheidungen notwendig, um den Nutzer optimal im Objektraum zu positionieren. Allerdings unterscheiden sich die Faktoren durchaus in ihrer Wichtigkeit, so dass bei geeigneter Reihenfolge der Entscheidungen schon nach sehr wenigen Schritten mit einer „ausreichend guten“ Positionierung zu rechnen ist. Um dies zu erreichen sollten sich die jeweils gewählten Dimensionen stark unterscheiden, das heißt die Filme sollten gemäß ihrer Ausprägungen der Faktoren möglichst distinkt sein. Hierzu wird für jeden Faktor ein Vektor erstellt, dessen Elemente den Ausprägungen jedes einzelnen Films für dieses Merkmal entsprechen. Nach erfolgter Berechnung der Ähnlichkeit dieser Vektoren (z.B. mittels des Kosinusmaßes) können Schritt für Schritt Faktoren gewählt werden, die zu allen vorherigen Faktoren den größtmöglichen Unterschied aufweisen. Anhand dieser gewählten Faktoren werden im Anschluss iterativ die Dialogschritte aufgebaut.

3.2 Auswahl geeigneter Repräsentanten

Um dem Nutzer möglichst gut unterscheidbare Mengen von Filmen präsentieren zu können, werden geeignete Repräsentanten nach den folgenden drei Kriterien ausgewählt:

- 1 *Bekanntheit*: Um sicherzustellen, dass der Nutzer auch qualifiziert über die Mengen urteilen kann, beschränken wir uns auf Filme mit einer Mindestanzahl an Bewertungen.
- 2 *Hohe Unterschiedlichkeit*: Die Filme werden dann derart ausgewählt, dass sie gemäß des aktuell betrachteten Faktors eine hohe Unterschiedlichkeit aufweisen.
- 3 *Isolierung des Faktors*: Ebenso wird darauf geachtet, dass die Filme hinsichtlich aller anderen Faktoren eine möglichst neutrale Position einnehmen.

4 Evaluation

In einer empirischen Studie wurde die Effektivität und Effizienz des vorgestellten Ansatzes im Vergleich zu manueller Exploration und vollautomatischen Empfehlungssystemen überprüft. Unsere Vermutung war, dass eine manuelle Navigation über Menüs, Suchfelder und Hyperlinks zwar gut passende Ergebnisse liefert (hohe Effektivität), aber aufgrund des hohen Interaktionsaufwands weniger effizient ist. Für die automatisch generierten Empfehlungen vermuteten wir gegenteilige Ergebnisse. Für unseren Ansatz, im Folgenden als „Interaktive Empfehlung“ bezeichnet, nahmen wir an, dass er einen guten Trade-off zwischen diesen Methoden darstellt. Zur Untersuchung dieser Vermutungen entwickelten wir ein fiktives Filmportal, welches eine Filmauswahl durch vier unterschiedliche Methoden zulässt:

- **Populäre Filme (Pop)** zeigt eine Liste von sechs Filmen an, die nach einer Kombination von hohem Durchschnittsrating und hoher Anzahl abgegebener Bewertungen ermittelt wurde. Eine Personalisierung findet nicht statt.
- **Manuelle Exploration (Man)** erlaubt eine freie Interaktion über Navigationsmenüs, Suchfelder, Tags und Hyperlinks im Portal.
- **Automatische Empfehlung (Aut)** präsentiert sechs mittels Matrix Factorization automatisch generierte Empfehlungen. Hierzu müssen initiale Bewertungen einiger Filme durch den Nutzer vorliegen (Details folgen bei der Beschreibung des Studienaufbaus).

- **Interaktive Empfehlung (Int)** führt einen Dialogprozess wie oben beschrieben mit fünf Auswahlentscheidungen durch und präsentiert danach ebenfalls sechs Empfehlungen.

Zur Realisierung wurde die *FactorWiseMatrixFactorization* aus der *MyMediaLite*-Bibliothek (Gantner et al., 2011) verwendet. Als Bewertungsgrundlage wählten wir den häufig verwendeten MovieLens 10M-Datensatz¹. Um weitere Metadaten zur besseren Darstellung der Filme zu erlangen, reicherten wir diesen um Filmbeschreibungen, Tags, Bilder und weitere Informationen an. Hierzu bedienten wir uns des HetRec'11-Datensatzes (Candator et al., 2011) und importierten zusätzlich Informationen aus der Internet Movie Database² (IMDb).

4.1 Hypothesen

Zur Überprüfung unserer Vermutungen formulierten wir eine Reihe von Hypothesen: Die vier Methoden unterscheiden sich im Hinblick auf wahrgenommene Passung (**H1**) und wahrgenommenen Neuigkeitswert (**H2**) der Empfehlungen. Ebenso im Hinblick auf die Möglichkeit der Beeinflussung des Auswahlprozesses (**H3**) und den vom Nutzer wahrgenommenen Aufwand (**H4**). Die Empfehlungen werden als unterschiedlich adaptiv wahrgenommen (**H5**), und die Nutzer vertrauen in unterschiedlichem Maß darauf, dass sie die Nutzerinteressen und nicht die Absichten des Anbieters berücksichtigen (**H6**). Sie eignen sich unterschiedlich gut für Explorationen mit ungefähigem Suchziel (**H7**) sowie für Explorationen ohne konkretes Suchziel (**H8**). Unser Ansatz ist dabei in jedem der betrachteten acht Aspekte gegenüber mindestens zwei der Alternativen überlegen (**H9–H16**).

4.2 Studienaufbau

Zur Überprüfung der Hypothesen baten wir 14 Versuchsteilnehmer (je 7 m/w; Durchschnittsalter 34.5, σ 14.10) die vier Explorationsmethoden Pop, Man, Aut und Int im fiktiven Filmportal zu nutzen. Vorab bewerteten die Probanden auf einer Skala von 1–5 drei zufällig aus den zehn populärsten gewählte Filme. Diese Bewertungen dienten später als Basis für die Empfehlungen der automatischen Methode.

Abbildung 2 zeigt eine beispielhafte Gegenüberstellung von Filmen in einem Dialogschritt der interaktiven Methode. Während die linke Menge eher düstere Filme beinhaltet, sind es auf der rechten Seite eher leichtere Unterhaltungsfilme.

¹ <http://www.grouplens.org/node/12> (>10M Bewertungen und >95T Tags von >70T Nutzern für >10T Filme).

² <http://www.imdb.com>



Abbildung 2: Dialog mit der Gegenüberstellung von Filmen mit unterschiedlichen Ausprägungen eines Faktors

Um die in jedem Dialogschritt fällige Entscheidung des Nutzers für eine der beiden Mengen zu erleichtern, wurden einerseits zusätzliche Metainformationen herangezogen (neben dem Kinoplatz eine kurze, zusammenfassende Beschreibung des Filmes), andererseits trugen Tag Clouds zur Differenzierung bei, indem sie die mit den jeweiligen Filmen verknüpften Tags visualisieren (z.B. „scifi“, „violent“ vs. „comedy“, „disney“). Nach einigen Dialogschritten und somit der Exploration mehrerer Dimensionen wurden die geäußerten Präferenzen des Nutzers herangezogen, um bezüglich der betrachteten Faktoren ähnlich ausgeprägte Filme zu selektieren und diese dem Nutzer als Empfehlungen darzubieten.

Die Studie fand in einem Zeitraum von zwei Tagen unter Aufsicht eines Versuchsleiters statt. Genutzt wurde ein Desktop-PC mit 24-Zoll Display und ein gängiger Internetbrowser. Die Probanden wurden gebeten, die vier Explorationsverfahren nacheinander auszuführen. Die Reihenfolge wurde dabei über alle Probanden hinweg systematisch permutiert. Nach jedem Durchlauf beurteilten die Probanden die jeweilige Methode anhand eines Fragebogens. Dieser enthielt eine Reihe von Aussagen mit einer bipolaren Skala mit 7 Zustimmungsgraden („Stimme gar nicht zu“ bis „Stimme voll zu“). Die Fragestellungen entsprachen dabei den in den Hypothesen H1–H8 adressierten Variablen. Im Einzelnen waren dies:

1. Die Auswahl passt sehr gut zu meinem Filmgeschmack (Passung).
2. Die Auswahl enthielt interessante Filme, die ich sonst vermutlich nicht gefunden hätte (Neuigkeitswert).
3. Ich hatte stets das Gefühl, den Auswahlprozess mitgestalten zu können (Einflussnahme).
4. Der erforderliche Aufwand, um eine Auswahl von Filmen zu erhalten, war akzeptabel (Aufwand).
5. Ich hatte das Gefühl, das System lernt meine Vorlieben (Adaptivität).
6. Ich habe Vertrauen, dass ausschließlich meine Bedürfnisse, nicht die Ziele des Verkäufers, berücksichtigt wurden (Vertrauen).
7. Ich würde die Methode nutzen, wenn ich bereits eine ungefähre Suchrichtung habe (mit Richtung).

8. Ich würde die Methode nutzen, wenn ich noch keine Suchrichtung vor Augen habe (ohne Richtung).

4.3 Ergebnisse

Mit Hilfe einer einfaktoriellen ANOVA mit Messwiederholung (Greenhouse-Geisser-adjustiert) stellten wir zwischen den vier Konditionen (hoch) signifikante Unterschiede fest. H1 bis H8 konnten dadurch bestätigt werden. Anschließend testeten wir post hoc unsere Methode Int paarweise mittels Bonferroni-Test auf Unterschiede gegenüber den alternativen Methoden, um Aufschluss über H9–H16 zu erlangen. Tabelle 1 zeigt die Ergebnisse der ANOVA und der anschließenden Vergleiche von Int mit den drei anderen Konditionen.

	Pop	Man	Aut	Int	F	Sig.
Passung	4.29, $\sigma=1.33$,	5.50, $\sigma=1.29$	3.36, $\sigma=1.39$	5.57, $\sigma=0.94$	10.353	p = .000
Int $\uparrow\downarrow$	\uparrow (r=.106)	\uparrow (r=1.000)	\uparrow (r=.002)*			
Neuigkeitswert	3.07, $\sigma=1.54$	2.86, $\sigma=1.75$	3.50, $\sigma=1.29$	5.50, $\sigma=1.23$	10.042	p = .000
Int $\uparrow\downarrow$	\uparrow (r=.004)*	\uparrow (r=.003)*	\uparrow (r=.002)*			
Einflussnahme	1.64, $\sigma=0.84$	5.79, $\sigma=1.25$	2.29, $\sigma=1.20$	5.71, $\sigma=0.83$	69.774	p = .000
Int $\uparrow\downarrow$	\uparrow (r=.000)*	\downarrow (r=1.000)	\uparrow (r=.000)*			
Aufwand	6.50, $\sigma=0.76$	2.07, $\sigma=0.83$	6.07, $\sigma=1.59$	6.00, $\sigma=0.88$	56.374	p = .000
Int $\uparrow\downarrow$	\downarrow (r=.662)	\uparrow (r=.000)*	\downarrow (r=1.000)			
Adaptivität	2.14, $\sigma=1.35$	1.21, $\sigma=0.58$	2.93, $\sigma=1.33$	5.79, $\sigma=1.05$	45.138	p = .000
Int $\uparrow\downarrow$	\uparrow (r=.000)*	\uparrow (r=.000)*	\uparrow (r=.000)*			
Vertrauen	2.64, $\sigma=1.45$	6.00, $\sigma=1.62$	3.43, $\sigma=1.40$	5.21, $\sigma=1.25$	15.990	p = .000
Int $\uparrow\downarrow$	\uparrow (r=.001)*	\downarrow (r=1.000)	\uparrow (r=.019)*			
Mit Richtung	2.43, $\sigma=1.70$	4.71, $\sigma=1.54$	2.14, $\sigma=1.10$	4.29, $\sigma=1.64$	10.911	p = .000
Int $\uparrow\downarrow$	\uparrow (r=.004)*	\downarrow (r=1.000)	\uparrow (r=.001)*			
Ohne Richt.	5.00, $\sigma=1.71$	2.64, $\sigma=1.45$	4.00, $\sigma=1.88$	5.93, $\sigma=1.07$	14.563	p = .000
Int $\uparrow\downarrow$	\uparrow (r=.760)	\uparrow (r=.000)*	\uparrow (r=.023)*			

Tabelle 1: Jede Zeile enthält Mittelwerte und Standardabweichungen für die jeweiligen Methoden. Zudem sind F-Wert nach Greenhouse-Geisser-Korrektur und der Signifikanzwert für die Überprüfung auf Within-Subject-Unterschiede angegeben (H1–H8). Die Pfeile geben schließlich an, ob Int besser oder schlechter als die jeweilige Vergleichsmethode bewertet wurde (H9–H16), und ob dieser Unterschied nach Bonferroni-Test signifikant ist (*).

Unser Ansatz wurde in Bezug auf Passung (H9) und Neuigkeitswert (H10) im Schnitt besser bewertet als alle Alternativen. In vier von sechs Vergleichen war der Unterschied signifikant. In Bezug auf Einflussnahme (H11), Vertrauen (H14) und Eignung bei vorhandenem Suchziel (H15) waren wir Pop und Aut signifikant überlegen. Kondition Man wurde hier erwartungsgemäß im Schnitt etwas besser bewertet, der Unterschied war jedoch nicht signifikant. Bezüglich des zu leistenden Aufwands (H12) wurde Int umgekehrt signifikant besser eingestuft als Man. Pop und Aut waren hier noch etwas besser, jedoch nicht signifikant besser. Hinsichtlich der zugeschriebenen Lernfähigkeit (H13) war Int allen anderen Methoden signifikant überlegen. Weiterhin wurde Int als überlegen eingestuft, falls noch kein Suchziel vorliegt (H16). Gegenüber zwei der drei Alternativen war der Unterschied signifikant.

5 Diskussion

Der vorgestellte Ansatz der interaktiven Empfehlungsgenerierung verbindet auf neuartige Weise die Vorteile algorithmischer Techniken mit denen der nutzergesteuerten Exploration und Suche. Es zeigt sich, dass die Nutzung latenter Faktoren ein vielversprechendes Mittel zur Generierung interaktiver Empfehlungsdialoge darstellt. Unsere Methode führt mit geringem Aufwand zu einer Menge von Empfehlungen, welche die Nutzer als ihren Interessen entsprechend einschätzen. Zudem erlaubt der schrittweise Explorationsprozess eine interaktive und situative Beeinflussung der Empfehlungsgenerierung, während bei rein automatischen Verfahren keine Steuerbarkeit durch den Nutzer besteht. Ein weiterer wesentlicher Vorteil der Methode liegt darin, dass das klassische Cold-Start-Problem umgangen wird, da es nicht notwendig ist, auf vorhandene Bewertungsdaten für einen Nutzer zurückzugreifen. Aufgrund der Auswahl von visualisierten Beispielen ist es weiterhin nicht erforderlich, dass die Nutzer ihre Bedürfnisse explizit ausdrücken können. Die für viele Domänen ohnehin schwer verfügbaren, expliziten Produktinformationen werden zudem obsolet.

In der durchgeführten Studie wurde die neue Methode in 16 von 24 möglichen Parametervergleichen mit den drei alternativ getesteten Methoden signifikant besser bewertet, was darauf hindeutet, dass die Methode hinsichtlich einer breiten Spanne relevanter Kriterien optimiert, d. h. einen guten Trade-off bezüglich unterschiedlicher Gestaltungsziele darstellt. Erwartungsgemäß zeigte sich, dass eine manuelle Exploration in Bezug auf Einflussnahme des Nutzers, Vertrauen und Eignung bei vorhandener Suchrichtung Vorteile bietet, wobei diese Unterschiede nicht signifikant waren. Umgekehrt wurde unser Ansatz in allen fünf anderen Aspekten, insbesondere auch beim empfundenen Aufwand besser bewertet (viermal signifikant). Der Aufwand wurde erwartungsgemäß bei den vollautomatischen Konditionen am besten bewertet. In allen anderen Aspekten war unser Ansatz den automatischen Methoden überlegen. Der Ansatz scheint also ein gutes Verhältnis von Aufwand zu Nutzen zu bieten. Er wird als lernfähig und vertrauenswürdig erachtet und scheint besonders dann geeignet, wenn der Nutzer noch kein Suchziel gebildet hat.

Hinsichtlich möglicher Einschränkungen und zukünftiger Untersuchungen ist zunächst anzumerken, dass die Anzahl der Teilnehmer an der durchgeführten Studie noch eher begrenzt war. Die Ergebnisse sollen dementsprechend in Folgeuntersuchungen erhärtet werden. In der Kondition automatische Empfehlungsgenerierung wurde lediglich eine mögliche Ausprägung untersucht. Es ist somit nicht auszuschließen, dass bei Verwendung umfangreicherer Trainingsdaten oder anderer Recommender-Algorithmen bessere Ergebnisse zu erzielen wären. Diese Option besteht allerdings prinzipiell auch für die interaktive Methode, die dadurch weiter optimiert werden könnte. Weitere Optimierungsmöglichkeiten bestehen zum Beispiel hinsichtlich der Auswahl und Visualisierung der auszuwählenden Objekte, der Zahl der Auswahlsschritte oder der Verbindung mit zusätzlichen Explorationstechniken. Zukünftig wollen wir weiterhin den Einfluss von Nutzercharakteristika wie z.B. der Vertrautheit mit der Produktdomäne oder den nutzerspezifischen Entscheidungsstrategien untersuchen.

Literaturverzeichnis

Bostandjiev, S., O'Donovan, J. & Höllerer, T. (2012). TasteWeights: A visual interactive hybrid recommender system. In *6th ACM Conference on Recommender Systems*. New York: ACM. 35–42.

- Cantador, I.; Brusilovsky, P. & Kuflik, T. (2011). 2nd workshop on information heterogeneity and fusion in recommender systems. In *5th ACM Conference on Recommender Systems*. ACM.
- Chen, L. & Pu, P. (2012). Critiquing-based recommenders: Survey and emerging trends. *User Modeling and User-Adapted Interaction*, 22(1–2), 125–150.
- Chi, E. (2004). Transient user profiling. In *Workshop on User Profiling (at CHI 2004)*. 521–523.
- Gantner, Z.; Rendle, S.; Freudenthaler, C. & Schmidt-Thieme, L. (2011). MyMediaLite: A free recommender system library. In *5th ACM Conference on Recommender Systems*. New York: ACM. 305–308.
- Gretarsson, B., O'Donovan, J., Bostandjiev, S., Hall, C. & Höllerer, T. (2010). Smallworlds: Visualizing social recommendations. *Computer Graphics Forum*, 29(3), 833–842.
- Iacobelli, F., Bimbaum, L. & Hammond, K. J. (2010). Tell me more, not just more of the same. In *14th International Conference on Intelligent User Interfaces*. New York: ACM. 81–90.
- Konstan, J. A. & Riedl, J. (2012). Recommender systems: From algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1–2), 101–123.
- Koren, Y., Bell, R. & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8), 30–37.
- Mahmood, T. & Ricci, F. (2009). Improving recommender systems with adaptive conversational strategies. In *20th ACM conference on Hypertext and Hypermedia*. New York: ACM. 73–82.
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. London: Penguin Press.
- Ricci, F., Rokach, L., Shapira, B. & Kantor, P. (Hrsg.). (2010). *Recommender Systems Handbook*. Berlin: Springer.
- Pu, P., Chen, L. & Hu, R. (2012). Evaluating recommender systems from the user's perspective: Survey of the state of the art. *User Modeling and User-Adapted Interaction*, 22(4–5), 317–355.
- Sinha, R. & Swearingen, K. (2002). The role of transparency in recommender systems. In *CHI '02: Extended Abstracts on Human Factors in Computing Systems*. New York: ACM. 830–831.
- Xiao, B. & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly*, 31(1), 137–209.
- Vig, J., Sen, S. & Riedl, J. (2011). Navigating the tag genome. In *16th International Conference on Intelligent User Interfaces*. New York: ACM. 93–102.