

Wirtschaftswissenschaftliches Forum der FOM

Kollaboratives Filtern in latenten Datenräumen

Entwurf und prototypische Realisierung eines
Empfehlungssystems unter Datenknappheit

Marc Feldmann

Wirtschaftswissenschaftliches Forum der FOM

Band 55

Marc Feldmann

Kollaboratives Filtern in latenten Datenräumen

Entwurf und prototypische Realisierung eines
Empfehlungssystems unter Datenknappheit

Shaker Verlag
Aachen 2018

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Herausgebende Institution ist die FOM Hochschule für Oekonomie & Management gemeinnützige Gesellschaft mbH

Copyright Shaker Verlag 2018

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 978-3-8440-6179-6

ISSN 2192-7855

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen
Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9
Internet: www.shaker.de • E-Mail: info@shaker.de

VORWORT DES HERAUSGEBERS

Die private FOM Hochschule für Oekonomie & Management versteht sich mit ihrem ausbildungs- und berufsbegleitenden Studienangebot im wirtschaftswissenschaftlichen Bereich als eine Ergänzung der deutschen Hochschullandschaft. Durch die Schaffung zielgruppenadäquater, attraktiver Studienbedingungen ermöglicht sie gleichzeitig den Beschäftigten viele Chancen zur Weiterentwicklung und den Unternehmen die Anpassung an die Anforderungen, die sich aus der demografischen Entwicklung und den gestiegenen Qualifikationsbedarfen ergeben.

Die 1991 auf Initiative von Wirtschaftsverbänden gegründete FOM arbeitet seit ihrem Bestehen eng mit Unternehmen und Verbänden zusammen und unternimmt mit der vorliegenden Schriftenreihe einen weiteren Schritt zur Verzahnung von Theorie und Praxis. Studierenden mit herausragenden Studienleistungen wird hierin ein Forum gegeben, der interessierten Fachöffentlichkeit empirische Ergebnisse, innovative Konzepte und fundierte Analysen im Zuge einer breiten Veröffentlichung ihrer Abschlussarbeiten mitzuteilen. Daneben finden exzellente Dissertationen von FOM Dozenten Eingang in die Schriftenreihe.

Unser herzlicher Dank gilt Herrn Prof. Dr. Alexander Holland und Herrn Prof. Dr. Rüdiger Buchkremer, die die Abschlussarbeit von Herrn Feldmann als Erst- bzw. Zweitgutachter betreut haben.

Die Arbeit greift das Problem der Informationsüberflutung auf, die sich aus der fortschreitenden Digitalisierung von Entscheidungsumfeldern ergibt und mit der sich Akteure bei der Lösung von Entscheidungsproblemen konfrontiert sehen. Die Arbeit rückt damit die Informationsphase des Entscheidungsprozesses in den Vordergrund und beleuchtet dazu die Problemursächlichkeit der beschränkten Rationalität von Entscheidern. Der besondere Beitrag der Arbeit besteht in ihrer methodischen Herangehensweise: Ein prototypisch umgesetzter automatisierter Informationsfilter – ein System aus der Klasse der Entscheidungsunterstützungssysteme – erlernt das Präferenzprofil eines Entscheiders anhand historisch gespeicherter Präferenzinformationen und nutzt dieses zur Vorfilterung der in digi-

talen Entscheidungsumgebungen typischerweise sehr großen Mengen von Entscheidungsalternativen. Auf diese Weise können die Kosten der Entscheidungsfindung vom Entscheider an den Filteralgorithmus ausgelagert und das Kosten-Nutzen-Verhältnis der Entscheidungssituation merklich gesteigert werden. In diesem Sinne ist die Arbeit eine gelungene Anwendung einer informationstechnologischen Methode auf eine wirtschaftswissenschaftlich relevante Problemstellung und ein wertvoller Beitrag zum wirtschaftswissenschaftlichen Forum der FOM.

Wir hoffen, den vielfach regen und fruchtbaren Dialog zwischen Hochschule und Praxis mit dieser Reihe um eine weitere Facette zu bereichern. Als Herausgeber freuen wir uns, herausragende Leistungen unserer Studierenden durch eine Veröffentlichung würdig honorieren zu können.

Essen, im April 2018

Prof. Dr. Burghard Hermeier

Rektor

Prof. Dr. Thomas Heupel

Prorektor für Forschung

VORWORT DER GUTACHTER

Die Nutzung sozialer Medien durch Einsatzweise verschiedener Endgeräte und häufige Interaktion der Akteure ermöglicht die Generierung, Ablage, Auswertung und zielgerichtete Verwendung von Daten. Die Unterstützung durch Empfehlungsmechanismen in Produktauswahl- und Bewertungsprozessen ist dabei ein zentraler Aspekt. Neben Entscheidungsabläufen durch technische Entscheidungsunterstützung sind hierbei Empfehlungen für Auswahlprozesse von Interesse, die anwender- oder objektbezogen erfolgen und systemtechnisch als Empfehlungssysteme zum Einsatz kommen. Solche Systeme geben den Akteuren kontextbezogenen Teilmengen einer Entitätsmenge als nutzentrierte Empfehlungsbasis. Schwachstellen sind technisch der Umgang mit Datenknappheit in verfügbaren Datenstrukturen und die Kopplung von Präferenzdaten mit Entscheidungsmechanismen. Hier setzt die sehr gelungene Masterarbeit von Herrn Feldmann an. Der Aspekt Datenknappheit wird durch Verfahren zur Dimensionsreduktion und impliziten Verdichtung der Präferenzinformation in einer prototypischen Lösung praxisnah erprobt und fundiert durch Nutzung von Qualitätsmaßen reflektiert.

Das Thema der vorliegenden Arbeit, welche Herr Feldmann im Rahmen seines berufsbegleitenden Master-Studiums an der FOM Hochschule für Oekonomie & Management angefertigt hat, ist durch das realisierte Empfehlungsverfahren mit einer umfassenden Verfahrensreflexion sehr praxisnah und für weiterführende quantifizierbare Empfehlungsverfahren von besonderer Bedeutung.

Im ersten Schritt der Arbeit werden Begriffe und Mechanismen zu Empfehlungssystemen vorgetragen und der Forschungsstand mit Empfehlungsmechanismen als kollaborative und inhaltsbasierte Filterverfahren beschrieben. Herr Feldmann führt Probleme wie Datenknappheit auf und leitet sehr fundiert Lösungswege zur besseren Abbildung des Informationsgehalts ab. Im nächsten Schritt findet der Verfahrensauswahlprozess statt, wobei Nachbarschaftsverfahren priorisiert werden und als Basis zur Erstellung der Lösungsarchitektur in dimensionsreduzierten Räumen durch prototypische Begleitung dienen. Die Praxisrelevanz wird im Implementierungsschritt vorgetragen, wobei die Ergebnisse für das Collaborative Filtering repräsentiert werden.

Die vorliegende Ausarbeitung leistet insgesamt einen fundierten Beitrag zum Diskurs zwischen Theorie und Praxis in Entscheidungsabläufen für Empfehlungsmechanismen. Sie gibt Anregungen zu weiteren Forschungsthemen und Handlungsimpulsen für die Praxis.

Essen, im April 2018

Prof. Dr. Alexander Holland

Dekanatsbeauftragter
Wirtschaftsinformatik

Prof. Dr. Rüdiger Buchkremer

Stellv. Direktor ifid Institut für
IT-Management & Digitalisierung

INHALTSVERZEICHNIS

Inhaltsverzeichnis.....	V
Abkürzungsverzeichnis	VII
Abbildungsverzeichnis.....	VIII
Tabellenverzeichnis.....	IX
Symbolverzeichnis	X
1 Einführung.....	1
1.1 Forschungsdesign	5
1.2 Untersuchungsaufbau.....	8
2 Theoretische Grundlagen.....	11
2.1 Entscheidungstheorie	11
2.1.1 Elemente und Grundmodell	14
2.1.2 Nutzen.....	20
2.1.3 Beschränkte Rationalität.....	23
2.2 Empfehlungssysteme.....	26
2.2.1 Referenzarchitektur	28
2.2.2 Forschungsstand.....	31
2.2.3 Probleme.....	42
2.3 Reduzierte Singulärwertzerlegung	45
2.3.1 Matrixfaktorisierung	46
2.3.2 Geometrische Intuition in \mathbb{R}^2	47
2.3.3 Σ -Rangreduktion und Datensatzrekonstruktion.....	49
3 Verfahrensauswahl.....	51
3.1 Wissenschaftliche Methode	51
3.2 Nutzenskalierung und Kriteriengewichtung	53
3.3 Nutzwertermittlung.....	58
3.3.1 Clusteringverfahren	59
3.3.2 Entscheidungsbäume	62
3.3.3 Nachbarschaftsverfahren.....	64
3.3.4 Künstliche neuronale Netze.....	68
3.3.5 Regelbasen	70
3.3.6 Regressionsmodelle	73
3.4 Ergebnis der Nutzwertanalyse	74

4	Lösungsarchitektur	77
4.1	Modellierungsgrundsätze	78
4.2	Datenvorverarbeitung	80
4.3	Niedrigrangapproximation der Ratingmatrix	83
4.4	Nachbarschaftskonstruktion im latenten Raum	86
4.5	Empfehlungsmechanismus ϵ	87
5	Implementierung	91
5.1	Versuchsaufbau	91
5.2	Datensätze und Testsystem	96
5.3	Ergebnisse	98
5.4	Diskussion	103
6	Fazit	107
6.1	Zusammenfassung	107
6.2	Ausblick	110
	Literaturverzeichnis	113

ABKÜRZUNGSVERZEICHNIS

CV	Clusteringverfahren
DSS	Decision Support System
DT	Decision Tree
ExpVal	Expected Value Recommender
GRand	Gaussian Random Recommender
IBCF	Item-based Collaborative Filtering
J1.3	Jester Datensatz 3
KNN	Künstliches Neurales Netz
LSM	Latent Space Model
LSN	Latent Space Neighbors
ML1M	MovieLens 1 Million Datensatz
NN	Nearest Neighbors
NWA	Nutzwertanalyse
PCA	Principal Component Analysis
RB	Regelbasis
RM	Regressionsmodell
RSVD	Reduced Singular Value Decomposition
SVD	Singular Value Decomposition
UBCF	User-based Collaborative Filtering

ABBILDUNGSVERZEICHNIS

Abb. 1: Untersuchungsaufbau	8
Abb. 2: Bernoulli-Nutzenfunktion	20
Abb. 3: Vorschlag einer Referenzarchitektur für Empfehlungssysteme	29
Abb. 4: Singulärwertzerlegung	45
Abb. 5: Veranschaulichung der SVD am Einheitskreis	48
Abb. 6: Reduzierte Singulärwertzerlegung	49
Abb. 7: Preprocessing.....	81
Abb. 8: Semantische Verdichtung im dimensionsreduzierten Raum.....	83
Abb. 9: Low-Rank-Approximation.....	85
Abb. 10: Latent Space Neighborhood Construction.....	86
Abb. 11: Recommendation Engine ϵ	90
Abb. 12: MAE-Fehlergebirge der Grid Search über k und d	99
Abb. 13: MAE im Modellvergleich.....	103

TABELLENVERZEICHNIS

Tab. 1: Grundmodell Erwartungswertentscheidung	18
Tab. 2: Grundmodell Erwartungsnutzenentscheidung	22
Tab. 3: Gewichtung der ISO 9126-Nutzenkriterien	56
Tab. 4: Nutzwertermittlung	76
Tab. 5: Ähnlichkeitsmetriken für Nachbarschaftskonstruktion	93
Tab. 6: LSN-Optimierungsparameter	94
Tab. 7: Kerneltransformationen der NN-Ratinggewichte.....	96
Tab. 8: Testsystemspezifikationen	97
Tab. 9: Lage- und Streuungsmaße der Testdatensätze.....	97
Tab. 10: MAE in Abhängigkeit von der Ähnlichkeitsmetrik	100
Tab. 11: MAE in Abhängigkeit vom Stimmgewichtskernel	101

SYMBOLVERZEICHNIS

Mengen

I	Items
$I^\bullet \subseteq I$	gesehene Items
$I^o \subseteq I$	ungesehene Items
$I^* \subseteq I^o$	empfohlene Items
R	Bewertungsmetrik
U	User

Matrizen

\mathbb{I}	Item-Konzept-Faktormatrix
\mathbb{M}^T	Transponierte einer Matrix
\mathbb{M}'	Approximierte einer Matrix
\mathbb{R}	Ratingmatrix
Σ	Singulärwertmatrix
\mathbb{U}	User-Konzept-Faktormatrix

Variablen, Konstanten, Funktionen

$\ \cdot\ _F$	Frobenius-Norm
$E(\cdot)$	Erwartungswert
ε	Empfehlungsmechanismus
κ	Kompressionsgewinn
r	Bewertung, Präferenzausdruck
ϕ	Schätzer, Schätzfunktion, Präferenzschätzer
θ	Selektor, Selektionsfunktion
σ	Singulärwert
sim	Ähnlichkeitsmetrik
wkern	Stimmgewichtskern