

planung & analyse

Zeitschrift für Marktforschung und Marketing www.planung-analyse.de

3/2012 D11700F

Preisforschung

Neue Kanäle

planung &
analyse startet
in Social Media

Neue Zugänge

Special zu
Online-Panels
im Wandel

Neuer Algorithmus

So werden
TURF-Analysen
handhabbar



TURF auf der Überholspur

Sortimentsanalyse in neuen Geschwindigkeitsdimensionen

► Die Autoren



Torsten Bausch, Diplom-Mathematiker, ist Softwareentwickler für Multivariate Analysen bei IfaD, Institut für angewandte Datenanalyse, Hamburg. Der Schwerpunkt seiner Tätigkeit ist die Konzeption, Erweiterung und Implementierung statistischer Verfahren. Zuvor war er verantwortlich für die Entwicklung von Optimierungsverfahren für technische Anwendungen.

✉ tbausch@ifad.de



Dr. Heiko Schimmelpfennig, Diplom-Kaufmann, ist Projektleiter für Multivariate Analysen bei IfaD, Institut für angewandte Datenanalyse, Hamburg. Er ist schwerpunktmäßig für die methodische Beratung, Anwendung der Analyseverfahren und Schulungen verantwortlich. Zuvor war er als wissenschaftlicher Mitarbeiter und Dozent an verschiedenen Hochschulen und Universitäten tätig.

✉ hschimmelpfennig@ifad.de

Die optimale Zusammenstellung von Produktportfolios oder Sortimenten gehört seit jeher zu den Aufgaben der Marktforschung. Die Anzahl der verfügbaren Produkte für ein Sortiment steigt, Regalplatz, Produktionskapazitäten oder Kosten zwingen aber zur Beschränkung. Mit wachsender Produktvielfalt wird diese Aufgabe immer bedeutender, aber auch deutlich anspruchsvoller – nicht nur inhaltlich, sondern auch mathematisch. Die TURF-Analyse ist ein etabliertes Verfahren zur Ermittlung optimaler Produktkombinationen. Allerdings stoßen TURF-Analysen bisher sehr schnell an Grenzen der Rechengeschwindigkeit heute verfügbarer Computer. Torsten Bausch und Dr. Heiko Schimmelpfennig zeigen die Leistungsfähigkeit eines neuen Algorithmus auf, mit dessen Hilfe sich in einem Bruchteil des Rechenaufwandes gleichwertige Ergebnisse erzielen lassen.

Das Ziel der TURF-Analyse (Total Unduplicated Reach and Frequency) ist die Bestimmung der Produktkombinationen mit den höchsten Nettoreichweiten oder Frequenzen. Sie berechnet Nettoreichweiten und Frequenzen von Produktkombinationen. Bei etwas anspruchsvolleren, aber realistischen Aufgaben benötigen TURF-Analysen allerdings schnell astronomische Rechenzeiten für eine optimale Lösung. Steht wenig Zeit zur Verfügung, liefern herkömmliche Ansätze häufig unbrauchbare Ergebnisse.

Grundlegende Begriffe

Zunächst sollen die beiden grundlegenden Begriffe definiert werden, die im Rahmen von Produktlinien- und Sortimentsoptimierungen mit TURF-Analysen essentiell sind:

■ **Nettoreichweite:** Anzahl der Personen in der Stichprobe, für die zumindest eine der angebotenen Produktvarianten einer Produktlinie zum Kauf in Frage käme. In die Nettoreichweite geht jede Person also nur einmal ein, egal wie viele Produktvarianten sie kaufen würde.

■ **Frequenz:** Anzahl der Produkte einer Produktlinie, die für eine Person zum Kauf in Frage kämen, summiert über alle Personen der Stichprobe. Es werden also Doppel- und Mehrfachauswahlen einer Person berücksichtigt.

Bestückung eines Snackautomaten

Ein Beispiel soll die Einsatzmöglichkeiten von TURF-Analysen in der Marktforschung und die

Anforderungen an die Rechengeschwindigkeit verdeutlichen. Für einen Snackautomaten am Bahnhof kann der Betreiber aus 100 Produkten auswählen. Der Automat hat Platz für 30 verschiedene Snacks. Die Aufgabe besteht nun darin, diejenigen Produkte auszuwählen, die möglichst viele verschiedene Personen ansprechen und damit den höchstmöglichen Absatz bringen (Nettoreichweite). Basis für die Analyse ist eine Befragung potentieller Kunden mit einer skalierten Abfrage zur Kaufwahrscheinlichkeit der 100 verfügbaren Produkte. Nun reicht es nicht, den Automaten mit den 30 Produkten mit der höchsten durchschnittlichen Kaufwahrscheinlichkeit zu bestücken. Damit würden unter Umständen einige Personen mehrere kaufbare Produkte finden, aber nur eines kaufen wollen. Andere würden dagegen gar kein kaufbares Produkt finden. Es gilt also diejenige Produktkombination zu bestimmen, die für möglichst viele Personen ein kaufbares Produkt bereithält. Für diesen Snackautomaten gibt es etwa 29.000.000.000.000.000.000.000.000 (29 Quadrillionen) Alternativen, ein Set mit 30 Produkten zusammenzustellen. Daraus die besten Sets zu finden, war bisher eine unlösbare Aufgabe. Die Rechenzeit für die Ermittlung der 100 Sets mit der besten Nettoreichweite beträgt mit herkömmlichen Algorithmen (vollständige Suche) geschätzte 330 Milliarden Jahre.

Aufgabe	Set-Größe (Anzahl Produkte in Kombination)	Anzahl alternativer Produkte	Anzahl Kombinationen (ca.)	Rechenzeit vollständiger Algorithmus (P-STAT)	Rechenzeit genPLUS
Testaufgabe	12	49	92 Milliarden	9 Stunden	2,5 Minuten
Ermittlung eines optimalen Produktportfolios	10	76	1 Billiarde	4,5 Tage	4,5 Minuten
Produktlinien- Optimierung mittels Conjoint	5	5.760	53 Billionen	600 Jahre	3,5 Stunden
Bestückung eines Automaten	30	100	29 Quadrillionen	330 Milliarden Jahre	16,5 Stunden

© Bausch/Schimmelpfennig; planung & analyse 3/12

Abbildung 1: Übersicht der Leistungsfähigkeit hinsichtlich Geschwindigkeit

Bislang bekannte Algorithmen für TURF-Analysen

Die bislang bekannten und in der Praxis verwendeten Algorithmen zur TURF-Analyse lassen sich in drei verschiedene Klassen einteilen:

■ **Vollständig:** Es werden Nettoreichweiten oder Frequenzen für alle Kombinationsmöglichkeiten für die gewünschte Set-Größe von Produkten berechnet. Die Vorgehensweise ist einfach und es wird eine zu 100 Prozent richtige Lösung gefunden. Mit der Größe der Aufgabenstellung steigt die Rechenzeit allerdings sehr stark überproportional an.

■ **Heuristisch:** Es werden nicht alle Kombinationsmöglichkeiten durchgerechnet, sondern der Rechenprozess unter verschiedenen Annahmen abgekürzt. Diese Verfahren finden häufig viele gute Kombinationen in kurzer Zeit. Gerade bei größeren Aufgabenstellungen, bei denen die Zeitersparnis relevant wird, sind die Ergebnisse aber nicht verlässlich.

■ **Direkt:** Ebenso wie beim heuristischen Ansatz werden nicht alle Kombinationen durchlaufen. Allerdings verwerfen diese Ansätze nur Kombinationen, die aufgrund eindeutiger Kriterien nicht zu den besten Kombinationen gehören können. Dadurch sparen auch sie erheblichen Zeitaufwand ein, bei gleicher Zuverlässigkeit wie die vollständigen Suchläufe. Während mit diesen Verfahren Kombinationen mit der höchsten Frequenz schnell ermittelt werden können, steigt bei größeren Aufgabenstellungen die benötigte Rechenzeit zur Bestimmung der Kombinationen mit der größten Nettoreichweite ebenfalls sehr häufig in praxisferne Regionen.

Ein neuer Algorithmus

Um die Schwächen bisheriger Algorithmen zu überwinden, gleichzeitig aber die jeweiligen Vorzüge nutzbar zu machen, wurde der neue genPLUS Algorithmus zur Ermittlung der be-

züglich der Reichweite optimalen Produktkombinationen entwickelt. Er verknüpft ein Verfahren aus der genetischen Optimierung mit der heuristischen Suche. Die heuristische Suche zeichnet sich durch sehr hohe Geschwindigkeit aus, wobei allerdings im Allgemeinen nicht alle besten Kombinationen einer vorgegebenen Set-Größe gefunden werden. Zudem findet dieser Ansatz häufig nicht die geforderte Anzahl an Kombinationen. Genetische Verfahren liefern im Gegensatz zu der heuristischen Suche bei weitem bessere Ergebnisse und auch eine ausreichende Zahl von Kombinationen, brauchen dafür aber eine längere Rechenzeit. Durch die Integration der beiden Verfahren werden ihre Stärken gebündelt und die Nachteile der einzelnen Ansätze vermieden. In extrem kurzer Zeit können damit nahezu 100 Prozent richtige Lösungen gefunden werden. Geringfügige Abweichungen in den gefundenen Kombinationen gegenüber der vollständigen Suche ergeben sich auf hinteren Rängen mit Kombinationen schlechterer Reichweite und sind damit für den praktischen Einsatz nicht relevant.

Validierung des Algorithmus

Eine verlässliche Validierung des neuen Algorithmus ist naturgemäß nur anhand von Aufgabenstellungen möglich, bei denen durch die Anwendung des vollständigen oder direkten Algorithmus eine sichere Analyse aller Kombinationsmöglichkeiten in einem vertretbaren zeitlichen Rahmen zu realisieren ist. Für alle Testaufgaben, die unter diesen Bedingungen durchgeführt wurden, konnte der neue Algorithmus nahezu alle optimalen Produktkombinationen ermitteln. Produktkombinationen mit der besten Reichweite wurden immer gefunden. Dabei wurden direkte Suchläufe mit über 60 Tagen Laufzeit in die Validierung einbezogen. Die umfangreichste Validierung beruhte auf einer Suche nach den hinsichtlich

► **Kurzfassung** Größte Herausforderung von TURF-Analysen ist die durch die Kombinatorik bedingte oftmals nicht realisierbare Rechenlaufzeit. Der innovative genPLUS Algorithmus stellt eine für die Marktforschungs-Praxis bedeutende Optimierung der TURF-Analyse dar. Der neue Algorithmus macht viele TURF-Analysen überhaupt erst möglich und eröffnet damit Lösungsmöglichkeiten für bisher unlösbare Aufgabenstellungen.

► **Abstract** The key challenge of TURF-analysis is the exponential duration of computation due to combinations, thus making it unfeasible. The innovative genPLUS algorithm marks a significant optimization of the TURF-analysis in the market research praxis. The new algorithm offers many TURF analyses in the first place and therefore opens up possible solutions for so far unresolved tasks.

Rang	Kombination	Reichweite (abs.)	Reichweite (%)	Frequenz (abs.)
1	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 91 93 96 97	440	88,0	2180
2	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 84 91 97 100	440	88,0	2176
3	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 91 96 97 100	440	88,0	2175
4	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 77 81 91 97 100	440	88,0	2172
5	3 12 15 26 28 29 33 34 35 37 38 42 46 70 71 72 81 91 97 100	440	88,0	2170
6	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 91 93 96 98	440	88,0	2167
7	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 85 91 97 99	440	88,0	2165
8	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 84 87 91 97	440	88,0	2160
9	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 87 91 96 97	440	88,0	2159
10	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 85 91 97 100	440	88,0	2156
10	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 77 81 87 91 97	440	88,0	2156

© Bausch/Schimmelpfennig; planung & analyse 3/12

Abbildung 2: Die Kombinationen von 20 Produkten mit der höchsten Nettreichweite

planung & analyse

Reichweite besten Sets aus 10 Items aus einer Gesamtheit von 200 Items. Daraus resultieren über 22 Milliarden mögliche Kombinationen. Die 1.000 besten davon waren auf Grundlage eines Datensatzes mit 1.000 Fällen zu ermitteln. Die Stabilität der Ergebnisse aller Validierungsstudien und die Struktur des Algorithmus lassen darüber hinaus die Annahme zu, dass die optimalen Produktkombinationen auch jenseits dieser Grenzen verlässlich gefunden werden.

Rechen-Geschwindigkeit

Der Algorithmus genPLUS ermöglicht die Lösung von Aufgaben, von denen man vorher Jahre entfernt war. Abbildung 1 zeigt einige Beispiele für die Zeitersparnis. Die Rechenzeiten für den vollständigen Algorithmus im Bereich mehrerer Jahre sind auf Basis kleinerer Aufgabenstellungen im Bereich von bis zu 5 Tagen hochgerechnet. Gefordert sind jeweils die 100 Kombinationen mit den höchsten Reichweiten.

Produktportfolio-Entwicklung auf Basis einer Conjoint-Analyse

Durch die außerordentliche Geschwindigkeit des neuen Algorithmus eröffnen sich viele neue Einsatzmöglichkeiten für TURF-Analysen. Eine Anwendungsmöglichkeit besteht in der praktischen Umsetzung von Ergebnissen aus Conjoint-Analysen zur Entwicklung eines optimalen Produktportfolios.

Bei Conjoint-Analysen ergeben sich schon aus wenigen Produktmerkmalen und Ausprägungen

Beurteilungen für eine sehr große Anzahl unterschiedlicher Produktvarianten. Die Kaufbereitschaft kann zusätzlich im Anschluss an das Conjoint für Produkte mit unterschiedlicher Wertigkeit abgefragt werden. Daraus kann die Kaufbereitschaft für alle Produkte anhand ihrer Conjoint-Nutzenwerte errechnet werden. Für ein einfaches Conjoint mit 7 Merkmalen und insgesamt 25 Ausprägungen ergeben sich schon 5.760 mögliche Produkte mit ihren Kaufwahrscheinlichkeiten. Daraus ein Portfolio mit 5 Produkten zu ermitteln, das möglichst viele verschiedene Käufer anspricht, erfordert die Prüfung von 53 Milliarden Kombinationsmöglichkeiten. Eine Lösung mit herkömmlichen Methoden benötigt etwa 600 Jahre. Mit dem neuen Algorithmus ist die Aufgabe hingegen in 3,5 Stunden gelöst.

Der Algorithmus im Praxistest

Ein Automatenbetreiber überlegt, ob er Automaten größerer Kapazität aufstellen und wie er diese dann bestücken soll. Zurzeit besitzt er Automaten mit einer Kapazität von 15 Produkten. Zur Unterstützung seiner Entscheidung steht eine Datei zur Verfügung, die von 500 Befragten die Angaben zu 100 Produkten enthält, ob sie diese an einem Automaten kaufen würden (1) oder nicht (0).

Da die Automaten derzeit maximal 15 Produkte fassen können, ist als Referenz eine TURF-Analyse mit einer Set-Größe von 15 durchzuführen. Um zu überprüfen, inwieweit sich größere Automaten lohnen, sind für höhere Kapazitäten weitere TURF-Analysen nötig.

► Literatur

Adler, T.; Smith, C.; Dumont, J.: Optimizing Product Portfolios Using Discrete Choice Modeling and TURF. International Choice Modelling Conference, Harrogate, 2009

Beasley, D.; Bull, D. R.; Martin, R. R.: An overview of Genetic Algorithms: Part 1, Fundamentals. University Computing, 1993, 15(2), S. 58-69.

Cohen, E.: TURF analysis. Quirk's Marketing Research Review. July/August 1993, S. 10-13.

Miaoulis, G.; Free, V.; Parsons, H.: Turf: A New Planning Approach for Product Line Extensions. In: Marketing Research, March 1990, S. 28-40

Rechenberg, I.: Evolutionsstrategie – Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Friedrich Frommann Verlag, 1973

Schwefel, H.-P.: Evolution and Optimum Seeking. John Wiley & Sons, 1995

Rang	Kombination	Reichweite signifikant höher zur besten Kombination der Set-Größe ...					
		15	16	17	18	19	20
1	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 91 93 96 97	x	-	-	-	-	

© Bausch/Schimmelpfennig; planung & analyse 3/12

Abbildung 3: Ergebnisse der Signifikanztests

Beispielhaft werden Set-Größen bis 20 in die Untersuchung einbezogen. Entscheidungsgrundlage und damit Optimierungskriterium soll die Nettoreichweite sein. Den Automatenbetreiber interessieren jeweils die 10 besten Kombinationen.

Eine vollständige Suche würde bei dem gegebenen Umfang der Aufgabenstellung viele Jahre in Anspruch nehmen und ist somit nicht praktikabel. Mit der Software, in der der genPLUS-Algorithmus implementiert wurde, ist die Aufgabe hingegen problemlos lösbar. Sie benötigt für alle sechs TURF-Analysen zusammen 5,75 Stunden.

In Abbildung 2 sind die besten Kombinationen von 20 Produkten mit der höchsten Nettoreichweite dargestellt. Die auf dem ersten Rang enthält die Produkte mit den Kodierungen 3, 12, 15, etc. Mit dieser wird die maximale Reichweite von 440 erzielt. Das heißt für 440 der 500, also 88 Prozent der Befragten ist unter diesen 20 Produkten mindestens eins dabei, das sie kaufen würden. Für diese Kombination ergibt sich eine Frequenz von 2180. Die folgenden Kombinationen unterscheiden sich nicht in der Reichweite, aber in der Frequenz. Die letzten beiden sind bezüglich beider Werte identisch, so dass sie auf demselben Rang liegen und insofern nicht nur die 10 besten, sondern die 11 besten Kombinationen ausgegeben werden.

Die Software bietet zudem die Option, Signifikanztests durchzuführen. Zum einen besteht dadurch die Möglichkeit zu untersuchen, ob sich die (entsprechend dem gewählten Optimierungskriterium) maximalen Reichweiten bzw. Frequenzen zwischen den unterschiedlichen Set-Größen unterscheiden. Zum anderen können Signifikanztests innerhalb einer Set-Größe durchgeführt werden. Interessant ist dies, falls die beste Kombination nicht in Frage kommt, man aber eine Kombination gleicher Größe mit zumindest nicht signifikant geringerer Reichweite bzw. Frequenz bestimmen möchte. Für die beschriebene Fragestellung ist der erste Signifikanztest bezüglich der Reichweite relevant.

In Abbildung 3 bedeutet ein „x“, dass die Reichweite der besten Kombination einer Set-Größe signifikant höher ist als die Reichweite der besten Kombination einer anderen Set-Größe. In diesem Beispiel ist demnach bei optimaler Bestückung die maximale Reichweite eines Automaten mit einer Kapazität von 20 Produkten signifikant ($p < 0.01$) höher als die eines Automaten mit 15 Produkten, nicht aber signifikant höher als die eines Automaten mit 16 oder mehr Produkten. Das heißt im Sinne einer signifikanten Verbesserung der Nettoreichweite würde sich ein Wechsel von Automaten mit einer Kapazität von 20 Produkte lohnen. ◀

„Die Inklinaton des homo sapiens zur Falsifikation einer per se richtigen Hypothese ist ein diesem inhärentes und entsprechend mit großer Inzidenz auftretendes negatives Leistungskriterium.“

Dr. Sabine Lang
Marktforschung

qualitative Forschung

meinungsforschung

marktforschung

unverständliche Reports?

repräsentativbefragungen

kundenbefragungen

produkttests

mitarbeiterbefragungen

Es geht auch anders: Irren ist menschlich.



Der Artikel hat Ihnen gefallen und Sie wollen mehr lesen?

planung & analyse ist eine der führenden Fachzeitschriften für Marktforschung und Marketing und veröffentlicht **acht Mal im Jahr** praxisorientierte, aktuelle und fundierte Fachbeiträge zu Projekten, Konzepten und Methoden der Markt- und Marketingforschung in sechs deutschen und zwei internationalen, englischsprachigen Ausgaben.

Einzelartikel bei planung & analyse

Bestellen Sie diesen und weitere Artikel in druckfähiger Version für nur **EUR 9,10/Artikel**

Abonnement von planung & analyse

- Testen Sie planung & analyse im **Probeabonnement** (Inland):
3 Hefte für **EUR 85,00** (inkl. Versand und MwSt.)
- Oder abonnieren Sie planung & analyse direkt im **Jahresabonnement** (Inland):
8 Hefte (6x deutsch, 2x englisch) **EUR 199,00** (inkl. Versand und MwSt.)

Alle weiteren Informationen zu Bezugsbedingungen im In- und Ausland, Inhalte & Abstracts der bereits erschienenen Ausgaben sowie Themen und Termine 2013 finden Sie unter www.planung-analyse.de.

Fragen oder Anregungen? Ihr direkter Kontakt zu planung & analyse:
+49 (0)69 7595 2014; info@planung-analyse.de

**planung
&analyse**