

DeltaMaster clicks!

06/2005

Liebe Datenanalysten,

hätten Sie gedacht, dass man mit der Warenkorbanalyse nicht nur Warenkörbe analysiert? Der Grundgedanke solcher Verfahren hat ihnen zwar speziell im Handelssektor einige Popularität beschert, er lässt sich aber auch auf ganz andere Branchen und Unternehmensbereiche übertragen. Allgemeiner spricht man von Assoziationsanalysen und meint damit Ansätze aus dem Data Mining, die automatisch Abhängigkeiten zwischen Merkmalen identifizieren und bewerten.

Dennoch ist die Metapher des Warenkorbs mitunter durchaus hilfreich: der Korb als eine zusammengehörende Menge und die im Korb liegenden Artikel als die Eigenschaften, deren gemeinsames Auftreten oder Vorkommen man erforschen möchte. In diesem Sinne assoziieren wir „Warenkorb“ also eher mit einem Merkmalsbündel und nicht unbedingt mit einem wahren Korb oder mit Korbwaren...

Mit *DeltaMaster* können Sie Assoziationsanalysen besonders einfach nutzen, denn das entsprechende Modul ist vollständig in die Suite integriert. In den vorliegenden *clicks!* möchten wir Ihnen einen kleinen Überblick über die Assoziationsanalyse geben und Ihren Blick für mögliche Anwendungsfälle schärfen.

Herzliche Grüße

Ihr Team von Bissantz & Company



Jupiter und Erde Ein wunderbares Beispiel dafür, dass Daten nur dann zur Information werden, wenn ein Vergleich möglich ist. Die übliche Berichterstattung zeigt den Jupiter regelmäßig maßstabslos.

Termine

DeltaMaster@Work, der kostenlose Workshop zum Kennenlernen unserer Lösungen, findet das nächste Mal am 30. Juni 2005 bei uns in Nürnberg statt. Bitte melden Sie sich bei unserem Herrn Liepins an: liepins@bissantz.de

DeltaMaster 5.1.2

Eine neue *DeltaMaster*-Version steht zum Download bereit, unter anderem mit Kopierfunktionen im Kombinations-Cockpit und Formatierungsoptionen für numerische Spalten im relationalen Modell.

Interner Zinsfuß

Mit unserem neuen Add-in für MS Analysis Services (AS) lässt sich die Berechnung des internen Zinsfußes (Internal Rate of Return, IRR) sehr einfach und hoch performant durchführen. Das *BC_IRR* genannte Modul stellt das Verfahren als benutzerdefinierte MDX-Funktion in *DeltaMaster* und allen anderen Anwendungen, die auf Analysis Services basieren, bereit. Fragen Sie Ihre Ansprechpartner nach *BC_IRR* oder schreiben Sie an service@bissantz.de

Archiv

Frühere *DeltaMaster clicks!* finden Sie im Internet unter <http://www.bissantz.de/clicks>

Kniff des Monats Assoziationsanalysen nutzen

Die wohl bekannteste Spielart von Assoziationsanalysen findet sich im Handel: Welche Produkte erwirbt der Kunde zusammen in einem Kaufakt und welche nicht? Derlei komplementäre oder substitutive Beziehungen spürt die Warenkorbanalyse auf. Das *Assoziationsanalyse*-Modul in *DeltaMaster* ist aber nicht auf dieses eine Anwendungsszenario beschränkt, sondern eignet sich ganz allgemein immer dann, wenn Sie wissen möchten, ob manche Dinge systematisch im Verbund mit anderen geschehen, ob es Abhängigkeiten oder Wechselbeziehungen zwischen den Merkmalen Ihrer Analyseobjekte gibt, in welche Richtung sie wirken und wie stark.

Weitere Beispiele für Verbundeffekte, die man mit der Assoziationsanalyse untersuchen könnte, sind etwa

- die Ausstattungswünsche von PKW-Käufern oder die Hardware-Konfiguration von ausgelieferten Rechnern im Vertriebscontrolling,
- technische Merkmale, Problembeschreibungen und Fehlerursachen im Qualitätscontrolling sowie im Kundendienst,
- Symptome, Diagnosen und Therapien in der Medizin,
- soziodemografische Merkmale und Leistungsfälle in der Versicherungswirtschaft

und vieles mehr.

Als gedankliche Hinführung zu der Assoziationsanalyse dienen uns hier die Basisverfahren *Rangfolge* und *PowerSearch*.

Bei der *Rangfolge* sortiert *DeltaMaster* nach genau einem Kriterium, in genau einer Dimension, auf genau einer Ebene. Im Handel beispielsweise könnte man damit im Nu eine typische „Renner-Penner-Liste“ erstellen, indem man *obere* und *untere* Objekte für die Absatzmenge und für die Produktdimension auf der Ebene der Einzelprodukte anzeigt.

Meine Analyse für Sicht (8): Renner-/Pennerliste Produkte							
Berechnen				Ansicht Drill-in Speichern Einstellungen Filter			
Obere	Produkt	Anteil	Absatz	Untere	Produkt	Anteil	Absatz
1.	Precisio JK	19,0%	401.536	1.	Ergoplus Basic	0,0%	24,86
2.	EF Besucherst. MP	15,8%	334.346	2.	Discus Comp	0,0%	236
3.	Hansen 10	13,0%	275.723	3.	Discus Office	0,1%	1.703
4.	Arcade AE 44	8,8%	187.298	4.	Arcade AE 77	0,4%	8.379
5.	EF Drehst. Nova	7,1%	150.239	5.	Ergoplus Senso	0,5%	10.828
6.	Ergoplus Nova	6,1%	129.073	6.	Precisio LM	0,5%	11.614
7.	Precisio LF	5,8%	123.810	7.	EF Besucherst. MC	0,7%	15.440

Auch *PowerSearch* sucht die größten bzw. kleinsten Ausprägungen eines Analysewerts, aber über verschiedene Dimensionen und Ebenen hinweg. *DeltaMaster* kombiniert also die Dimensionen und Ebenen des gesamten Modells und fördert die bedeutendsten Werttreiber zu Tage. Man würde dann etwa leicht erkennen, wenn es Großkunden gibt, mit denen man einen höheren Umsatz erzielt als in einer ganzen Vertriebsregion. *PowerSearch* hat sich als ein guter Einstiegspunkt bewährt, wenn man es mit frischen Daten zu tun hat, denn man erkennt sehr schnell, „wie der Hase läuft“.

Im nächsten Schritt wollen wir nun wissen, wie sich die Merkmale in ihrer Kombination auf unser Geschäft auswirken. Zur Erläuterung benutzen wir einen Auszug aus einem Porsche-Modell.

In der zugrunde liegenden Datenbank sind Verkaufsaufträge über einige Porsche 911 mitsamt der gewählten Ausstattung gespeichert. Wenn man in den *Einstellungen* von *PowerSearch* die *Analysekomplexität* auf vier stellt, so erfährt man durch Regel 16, dass in rund 45 % der Fälle 18-Zoll-Felgen, naturlederne Innenausstattung, Lackierung in violettchromaflair (ein Farbton zwischen Aubergine und Anthrazit) und das Modell Millennium geordnet wurde. Ist diese Begebenheit ein Zufall? Oder gibt es systematische Beziehungen zwischen den Merkmalen?

Analyseset	Berechnen	Ansicht	Drill-In	Speichern	Einstellungen
Basisverfahren	Obere	Alle Dimensionen bis Ebene 2			Anteil
Statistische Verfahren	1.	Felgen: 18" Felgen			81,8%
Mining Verfahren	2.	Getriebe: 6 Gang			59,1%
PowerSearch	3.	Innenausst: Naturleder			50,0%
Navigation	4.	Felgen: 18" Felgen; Innenausst: Naturleder			50,0%
Comparator	5.	Lack: violettchromaflair			45,5%
Descriptor	6.	Modell: Millennium			45,5%
Profiler	7.	Felgen: 18" Felgen; Lack: violettchromaflair			45,5%
Interdependenzanalyse	8.	Felgen: 18" Felgen; Modell: Millennium			45,5%
Dependenzanalyse	9.	Innenausst: Naturleder; Lack: violettchromaflair			45,5%
Selector	10.	Innenausst: Naturleder; Modell: Millennium			45,5%
Assoziationsanalyse	11.	Lack: violettchromaflair; Modell: Millennium			45,5%
	12.	Felgen: 18" Felgen; Innenausst: Naturleder; Lack: violettchromaflair			45,5%
	13.	Felgen: 18" Felgen; Innenausst: Naturleder; Modell: Millennium			45,5%
	14.	Felgen: 18" Felgen; Lack: violettchromaflair; Modell: Millennium			45,5%
	15.	Innenausst: Naturleder; Lack: violettchromaflair; Modell: Millennium			45,5%
	16.	Felgen: 18" Felgen; Innenausst: Naturleder; Lack: violettchromaflair; Modell: Millennium			45,5%

Um dies zu klären, müssen wir in die *Assoziationsanalyse* wechseln. *PowerSearch* liefert uns zwar Aussagen über die Häufigkeit des Zusammentreffens von Merkmalen, aber nicht über Abhängigkeiten. Die Assoziationsanalyse ist in *Miner-Expert*-Versionen unter den Mining-Verfahren zu finden.

Das Ergebnis der Assoziationsanalyse sind Wenn-Dann-Regeln. Zu jeder gefundenen Regel berechnet *DeltaMaster* statistische Kennzahlen, zum Beispiel die Häufigkeit (auch Support oder Reichweite genannt), die abhängige Wahrscheinlichkeit (Confidence), den so genannten Lift (Improvement), das Relative Risiko, die Regelstärke, Chi-Quadrat-Werte und andere. Diese Maße beschreiben, wie stark der entdeckte Zusammenhang ist, und geben somit Auskunft über die „Qualität“ der Regel.

Meine Analyse für Sicht (1): Assoziation aller Merkmale *						
Berechnen						
Filter Wenn: Tupelgröße Felder Elemente Filter Dann: Tupelgröße Felder Elemente						
Regeltyp	Wenn (A)	Dann (B)	Häufigkeit (A U B)	abhäng. Wahrscheinlichkeit	Lift	
Sym. Sicher	(Millennium)	(18" Felgen, violettchromaflair, Naturleder)	45,45 %	100,00 %	2,20	
Sym. Sicher	(violettchromaflair)	(18" Felgen, Naturleder, Millennium)	45,45 %	100,00 %	2,20	
Sym. Sicher	(996 RS)	(6 Gang, weiß, Leder schwarz)	22,73 %	100,00 %	4,40	
Sym. Sicher	(weiß)	(6 Gang, Leder schwarz, 996 RS)	22,73 %	100,00 %	4,40	
Sicher	(silber)	(Carrera)	22,73 %	100,00 %	3,14	
Sicher	(rot)	(18" Felgen, Carrera)	9,09 %	100,00 %	7,33	
Sicher	(Leder grau)	(6 Gang, Carrera)	9,09 %	100,00 %	5,50	
Normal	(tiptronic)	(18" Felgen)	36,36 %	88,89 %	1,15	
Normal	(Leder schwarz)	(6 Gang)	27,27 %	66,67 %	1,13	

Gleich in der ersten Zeile finden wir die Merkmalskombination, die uns in *PowerSearch* aufgefallen war. Als „Wenn“-Kriterium ist das Modell Millennium eingetragen, die anderen drei Ausprägungen stehen im „Dann“-Teil der Regel. Zu lesen ist das etwa so: Wenn in einem Auftrag das Modell Millennium geordnet wird, dann kommen auch 18-Zoll-Felgen, die Violettchromaflair-Lackierung und Naturleder hinzu – und zwar immer: Die abhängige Wahrscheinlichkeit liegt bei 100 %. Weil *DeltaMaster* auch den Umkehrschluss als zutreffend erkannt hat, ist die Regel vom Typ *Symmetrisch sicher*: Sie gilt immer und in beiden Richtungen.

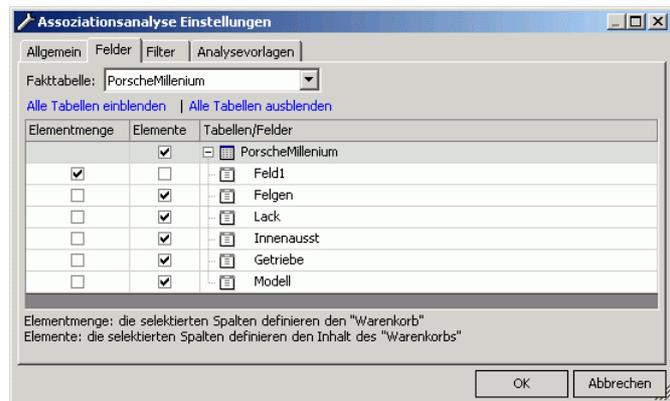
Zeile fünf verrät: Wenn der Kunde die Farbe Silber bestellt, dann ist es immer ein Carrera-Modell. Der Lift von 3,14 sagt aus, dass unter den silbernen Porsche die Häufigkeit der Carreras 3,14-mal höher ist als insgesamt.

Relationales Modell

Technisch gesehen fragt die Assoziationsanalyse, ob es in einer gegebenen Datenmenge Attributwerte gibt, die oft zusammen auftreten. Die Basisdaten entnimmt *DeltaMaster* einer relationalen Datenbank – die OLAP-typische Voraggregation in einem Cube bringt für diese Zwecke keine Beschleunigung.

Der Vorteil für Sie ist, dass Sie sehr schnell mit eigenen Experimenten starten können, denn für den Einstieg genügen eine Access-Datenbank, eine bestehende Tabelle im SQL-Server oder Ähnliches als Datenquelle. Auch die Formatierung spielt keine große Rolle.

DeltaMaster benötigt nur wenige Informationen, um Assoziationsanalysen auf relationalen Datenbanken rechnen zu können. Im Wesentlichen sind ihm die Namen von Tabellen und Tabellenspalten bekannt zu machen, welche die Elementmengen (den Warenkorb, sozusagen) und ihre Elemente (die Produkte, die im Korb liegen) beschreiben. Im Dialog *Einstellungen* wählen Sie aus, in welcher Tabelle und in welchen Spalten *DeltaMaster* diese Angaben findet. Alle Datenbank-Zeilen, die in den *Elementmenge*-Feldern identische Werte haben, bilden gemeinsam einen Fall, sodass Sie beliebig große Elementmengen analysieren können.



Wenn Sie neugierig auf die faszinierenden Möglichkeiten der Assoziationsanalyse geworden sind: Ihre Ansprechpartner beraten Sie wie immer sehr gerne!