

## ABRAHAM MEIDAN

ist Präsident und Geschäftsführer von WizSoft Inc., einem führenden Softwareunternehmen, im Umfeld innovativer Datenprüfungswerkzeuge. Er promovierte 1982 an der Universität Tel-Aviv, veröffentlichte mehrere Bücher und meldete zwei Patente zu dem aufgeführten Thema an. Hier erläutert er die Entwicklungs- und Einsatzprinzipien von WizRule sowie den Unterschied zu tradierter Prüfsoftware

### Prüfsoftware versus Data-Mining

Interne Revisoren, Wirtschaftsprüfer und forensische Ermittler beschäftigen sich unter anderem mit der Aufhellung von Fehlern und Betrugsfällen, die sich innerhalb betrieblicher Daten abbilden. Hierbei bedienen sie sich hauptsächlich tradierter Prüfsoftware, wie zum Beispiel ActiveData, ACL und IDEA, mit deren Hilfe Inkonsistenzen und „seltsame „Fallkonstellationen innerhalb dieser Daten aufgezeigt werden. Wesentliche Funktionen sind dabei das Filtern und Sortieren (CAAT). Dieses erlaubt es beispielsweise alle Transaktionen aufzulisten, deren Rabattsatz über einen bestimmten, erwarteten Schwellenwert liegt. Ein solches Vorgehen kann allerdings ausschließlich solche Fehler und Betrugsfälle aufhellen, die dem Prüfer, der seine prüferische Erfahrung auf die Daten abbildet, bekannt sind. In unserem aufgeführten Beispiel ist dieses dann der Fall, wenn der Verdacht besteht, dass zumindest Fehler wenn nicht einige betrügerische Rabatte innerhalb des Datenbestandes vorhanden sind. Vermutet der Revisor hingegen kein Problem im Zusammenhang mit der Rabattgewährung, bestehen kaum Aussichten, dass mit Hilfe von Prüfsoftware ein solcher Vorfall aufgedeckt wird. An dieser Stelle setzt Data-Mining-Software an, die als regelsuchendes Verfahren betriebliche Daten ohne solche Voraussetzungen weitgehend automatisch überprüft. Hierzu identifiziert sie strukturelle Beziehungen zwischen einzelnen Datenfeldern und Werten um anschließend auffällige Abweichungen abzuleiten.

### Wie funktionieren Data-Mining-Tools?

Üblicherweise suchen entsprechende Programme in Daten zurückliegender geschäftlicher Aktivitäten nach Mustern, Verknüpfungen und Beziehungen, um hieraus Vorhersagen für neue, vergleichbare Vorgänge abzuleiten. Dieses ermöglicht es zum Beispiel einer Bank Muster in Kundendaten zu iden-

tifizieren, welche Kreditausfälle verursacht haben. Bei neu beantragten Kundenkrediten werden diese Muster anschließend angewendet, um deren Ausfallwahrscheinlichkeit zu berechnen. Ein vergleichbarer Ansatz unterstützt auch die Aufdeckung von Betrugsfällen verwendet. So setzen beispielsweise Kreditkartenunternehmen Data-Mining-Software ein, um Muster bekannter Betrugsfälle für die Prüfung neuer Karten-Transaktionen einzusetzen. Für beide Ansätze sind jedoch vorhergehende Beispiele erforderlich.

Hier kann *WizRule* mit einem erweiterten Ansatz auf der Grundlage von Assoziationsregeltechnologien unterstützen. Anstatt ausschließlich Vorhersagen für neue Fälle zu treffen, werden hier Fälle ermittelt, die von gültigen Mustern abweichen. Grundlage ist die Annahme, dass Betrugskonstellationen in vielen Fällen Ausnahmen von vorgegebenen Regeln darstellen. Wenn beispielsweise Verkäufer ausschließlich für bestimmte Kunden zuständig sind und eine einzelne Transaktion von einem abweichenden Verkäufer abgewickelt wird, besteht ein Verdachtsfall, der untersucht werden sollte.

Der vorstehend aufgeführte Ansatz setzt voraus, dass *WizRule* zunächst alle Regeln (Muster) in einem zu analysierenden Datenbestand identifiziert. Dieses beinhaltet ein "Revers Engineering" von Regeln, welche die Daten erzeugt haben. Mathematische Algorithmen auf statistischer Grundlage zeigen alle Regeln eines Datenbestandes auf und analysieren nachfolgend Vorgänge, die in Bezug auf die entdeckten Regeln unwahrscheinlich sind. Diese repräsentieren die zu untersuchenden Fallkonstellationen oder Verdachtsfälle.

### Wie läuft die praktische Anwendung?

Für den Prüfer ist das *Verfahren denkbar einfach*. Bei vorgegebener Parametrisierung muss lediglich die zu auditierenden Daten auszuwählen. Alle restlichen Arbeiten erle-

digte die Software selbstständig. Innerhalb kurzer Zeit werden die Analyseergebnisse auf dem Bildschirm angezeigt.

Mit ein wenig mehr Erfahrung wird ein Prüfer später die Standard- Analyseparameter verfeinern. Hierzu zählen unter anderem die „minimale Wahrscheinlichkeit von "Wenn-Dann-Regeln" sowie die "minimale Anzahl von Fällen, die einer Regel zugrunde liegen". Weiterhin kann er abgrenzen, nach welche Arten von Regeln die Analyse unterstützen und ob einige Datenfelder hierbei ignoriert werden sollen. Interessant sind in diesem Zusammenhang die Algorithmen, nach welchen Regeln und Muster gesucht werden. Sie suchen und finden interessante Assoziationen und / oder Korrelationen zwischen beliebigen Datenfeldern. Dabei identifizieren sie Konstellationen, welche in einzelnen Datensätzen häufig gemeinsam auftreten (Assoziationsregel-Technologie). Grundlage sind "Wenn-Dann"-Aussagen. Sie werden eigenständig aus vorliegenden Daten berechnet und bestimmen, wie die Inhalte eines Feldes durch die Werte anderer Felder beeinflusst werden. Über die Bedingungen hinaus (dem "Wenn"-Teil) und dem resultierenden Ergebnis (dem "Dann"-Teil) bestimmen drei Angaben den Grad der hiermit verbundenen Unsicherheit, wie das nachfolgende Beispiel zeigt:

*Wenn* die Niederlassung KÖLN ist  
*und* die Rechnung 200-500 Euro beträgt  
*und* Herr KLEIN der Verkäufer ist

Dann liegt der Rabatt unter 10%

Regelwahrscheinlichkeit: 0,998

Die Regel existiert in 1.002 Datensätzen.

Signifikanzniveau: Fehlerwahrsch. < 0,001

Die erste Angabe (Wahrscheinlichkeit) wird als Konfidenzintervall für die Regel bezeichnet. In Wenn-Dann-Regeln benennt sie das Verhältnis zwischen der Anzahl der Sätze in denen die Bedingung (en) und das Ergebnis gelten sowie der Gesamtzahl von Sätzen innerhalb der aufgeführten Bedingungen, die mit oder ohne finales Ergebnis vorliegen.

Die zweite Angabe beinhaltet die Anzahl der Sätze, in welchen der Bedingung und Ergebnisteil zu treffen.

Das abschließend aufgeführte Signifikanzniveau lenkt den Blick auf die Gültigkeit der

Regel sie entspricht eins minus der Fehlerwahrscheinlichkeit, welche dafür steht, dass die Regel zufällig in den zu analysierenden Daten vorliegt.

### Verschiedene Arten von Regeln

Zur Ermittlung von Auffälligkeiten werden fünf Regelarten für die Analysen verwendet:

- Wenn-Dann-Regeln-Regeln (If-Then Rules)

Wie bereits in dem vorherigen Beispiel dargestellt.

- Formel-Regeln (Formular Rules)

$$A = B * C$$

Für: A = Summe

B = Menge

C = Stückpreis

Genauigkeitsgrad der Regel: 0,99

Die Regel existiert in 1.890 Datensätzen.

- Wenn-Dann-Regeln-Formel-Regeln (If-Then Formular Rules)

Wenn Kundenname ist Meier

Dann Summe = Stückpreis \* 0,8 + 50

Genauigkeitsgrad der Regel: 0,95

Die Regel existiert in 502 Datensätzen.

- Regel zur Schreibweise (Spelling-Rules)

Der Eintrag Meier ist 2.080 Mal im Feld Kundenname vorhanden.

Es gibt 2 Fälle mit ähnlichen (aber nicht exakten gleichen) Einträgen.

- Abweichende Regel (Outstanding Rules)

Wenn Kundenname ist Meier

Dann Rabattsatz = 20

Die Regel ist als abweichend einzustufen.

Begründung: Es gibt 100 Einträge im Feld Kundenname. Davon weist keiner weniger als die Mindestanzahl an Fällen in Bezug eine Regel auf. Es wurden keine vergleichbaren Regeln gefunden, die mit den anderen Einträgen im Feld Kundenname und den Werten im Feld Rabattsatz korrespondieren.

Während in Formel-Regeln arithmetische Zusammenhänge mit bis zu fünf Variablen, die in den Daten enthalten sind, ermittelt werden, können Wenn-Dann-Regeln für beliebig viele Bedingungen aufgehellt werden. Hier wie dort deuten Abweichungen von einer gültigen Regel auf Fehler oder möglicherweise Betrug hin.

### **Wie lassen sich Abweichungen beurteilen und Fehlalarme vermeiden?**

Im Anschluss an die automatische Entdeckung von Regeln werden eventuelle Abweichungen von diesen Regeln innerhalb des zu untersuchenden Datenbestandes durch die Algorithmen überprüft. Hierbei wird nicht jede Abweichung von einer Regel als Finding aufgenommen. Angenommen, folgende Wenn-Dann-Regel hat sich ergeben:

*Wenn* das Konto 52678 bebucht wird

Dann enthält Feld Umsatzsteuer 0,00 Euro

Regelwahrscheinlichkeit: 0,98

Die Regel existiert in 1.003 Datensätzen.

Signifikanzniveau: Fehlerwahrsch. < 0,001

In diesem Fall beträgt die Wahrscheinlichkeit der Regel 0,98 und sie existiert in 1003 Datensätzen, wobei 20 Datensätze von dieser Regel abweichen. Eine Überprüfung dieser Abweichungen wäre mühsam und ineffektiv, zumal wenn viele aus Fehlalarmen resultieren. Um solche Fehlalarmen zu vermeiden, erfolgt eine ergänzende Operation, bei der geprüft wird, ob sich die Abweichung durch eine weitere, in den Daten enthaltene Regel, erklärt. Bei einer positiven Antwort wird kein Verdachtsfall ausgewiesen.

Im vorliegenden Zusammenhang könnte dieses der Fall sein, wenn ein wechselndes Währungskennzeichen vorliegt oder das ein abweichendes Länderkennzeichen identifiziert wird. Da sich Abweichungen mit solchen Merkmalen möglicherweise erklären lassen, werden sie nicht als Untersuchungsfälle betrachtet.

Eine weitere Verifikation beschäftigt sich mit der Häufigkeit gültiger Werte in abweichenden Fällen. Ihre Häufigkeit sollte unter den Regelbedingungen niedriger sein als die Gesamthäufigkeit eines Vorliegens in den Daten. Ist dies nicht der Fall, so wird ein

Verdachtsfall ebenfalls ausgeschlossen. Wenn zum Beispiel innerhalb eines Regelbereichs Belege in zwei abweichenden Datensätzen mit einer abweichenden Belegart auf ein abweichendes Konto gebucht werden und dieses die beiden einzigen Fällen in der gesamten Datei sind, in welchen diese Belegart vorkommt, dann werden diese Fälle ebenfalls nicht als "zu untersuchen" ausgewiesen. Falls die Belegart jedoch "in der Regel" für andere Konten Verwendung findet, werden die aufgeführten Vorgänge als Verdachtsfälle klassifiziert.

Bei numerischen Dann-Feldern werden Fehlalarme weiterhin dadurch reduziert, dass ausschließlich Abweichungen angezeigt werden, bei welchen ein Ergebnis um mindestens eine Standardabweichung von erwarteten Werten abweicht. Kleinere Abweichungen werden ignoriert.

Innerhalb der Berichtserstellung können die angezeigten Verdachtsfälle sowohl nach dem Dann-Feld als auch nach den Werten innerhalb dieses Feldes sortiert werden. Dieses ist vor allem dann relevant, wenn der Bericht viele Abweichungen auflistet. Diese Sortierung ermöglicht es, sich auf interessante Abweichungen zu konzentrieren.

Alle aufgeführten Methoden tragen dazu bei, Fehlalarme weitgehend zu vermeiden, sodass sich eine nachfolgende Bearbeitung auf wirklich interessante Vorgänge konzentrieren kann.