

WHITE PAPER

WIZRULE – DATA MINING SOFTWARE FÜR PRÜFUNGSZWECKE

ABRAHAM MEIDAN, PH.D.

Eine der wesentlichen Aufgaben von Prüfern, Forensikern und Datenqualitätsmanagern ist es, betrügerische Fälle und Fehler in Daten aufzudecken. Die prüferische Arbeit kann dabei wesentlich durch *WizRule* unterstützt werden. *WizRule* ist eine intelligente Prüfsoftware, die auf Data Mining-Technologie basiert. Ziel ist es, Inkonsistenzen und Auffälligkeiten in Datenbeständen mit Hilfe von automatisierten Analysen aufzuzeigen.

Standardmäßig werden zum Aufdecken betrügerischer Fälle und Fehler konkrete Abfragen verwendet, die die Daten filtern und sortieren. Auf diese Weise werden beispielsweise alle Transaktionen gelistet, bei denen ein Kalkulationszinssatz über einem vordefinierten Grenzwert liegt. Es gibt verschiedene Prüfsoftware-Lösungen, wie beispielsweise ACL und IDEA, die eingesetzt werden können, um derartige Abfragen vorzunehmen. *WizRule* konkurriert mit diesen Lösungen nicht; es ergänzt sie vielmehr.

Grundsätzlich können die oben genannten Analysen nur Betrugsfälle und Fehler aufdecken, für die die zugrundeliegenden Abfragen vorgesehen wurden. Wenn ein Prüfer beispielsweise den Verdacht hat, dass einige Transaktionen des zu untersuchenden Datenbestands dolose Preisminderungen beinhalten, können darauf abzielende Datenabfragen präzisiert werden. Wenn jedoch lediglich ein Problem im Zusammenhang mit einem Preisnachlass vermutet wird, werden Abfragen mittels Prüfsoftware-Tools, wie Idea und ACL, in der Regel keine auffälligen Einträge feststellen. Da es eine große Vielzahl möglicher doloser Handlungsalternativen gibt, ist es weder effizient noch zielführend, für jede Möglichkeit separate Abfragen zu erstellen.

Hier kommt *WizRule* ins Spiel. *WizRule* generiert Ergebnisse automatisch auf Knopfdruck. Der Benutzer gibt lediglich die Auswahl an Daten an. *WizRule* führt selbständig die Analyse durch. Dabei überprüft es alle Beziehungen zwischen Werten in den verschiedenen Feldern und meldet unerwartete und unwahrscheinliche Fälle. Es deckt so die betrügerischen Fälle auf, die von den Standardprüfwerkzeugen übersehen werden.

DIE FUNKTIONSWEISE VON WIZRULE

WizRule ist eine Data Mining-Software. Data Mining-Programme haben zum Ziel, auffällige Muster in Daten aufzuzeigen.

Data Mining-Lösungen decken meist Muster in Datenbeständen auf, die in der Vergangenheit erfasst wurden. Auf dieser Basis generieren sie Vorhersagen für neue Fälle. Beispielsweise setzen Finanzdienstleistungsinstitute Data-Mining-Software mit dem Ziel ein, Muster in Kundendaten mit Forderungsausfällen aufzudecken. Sobald Kunden neue Kredite beantragen, kommen diese Muster zum Einsatz, um die Ausfallwahrscheinlichkeit des Kredits zu berechnen. Dieser Ansatz wird auch zur Aufdeckung von Betrug verwendet. Beispielsweise verwenden Kreditkartenunternehmen Data Mining-Programme, um Muster bereits bekannter betrügerischer Fälle zu erkennen und diese Muster bei der Überprüfung neuer Transaktionen anzuwenden. Dieser Ansatz kann nicht verwendet werden, wenn nach dolosen Handlungen ohne vorherige Mustervorgaben gesucht wird.

Hier kommt *WizRule* zum Einsatz. Anstatt Vorhersagen für neue Fälle zu treffen, zeigt *WizRule* Fälle auf, die von als gültig erkannten Mustern abweichen. Der methodische Ansatz von *WizRule* basiert auf folgender Annahme: *Betrug stellt in vielen Fällen die Ausnahme von der Regel dar*. Ein Beispiel soll diesen Sachverhalt verdeutlichen: In Zusammenhang mit einem bestimmten Kunden sind alle Verkaufsvorgänge mit einem bestimmten Verkäufer verbunden; es findet sich jedoch eine einzelne Transaktion, bei der der Verkäufer eine andere Person ist, die im Allgemeinen andere Kunden betreut. Hierbei handelt es sich um einen Verdachtsfall, der weiter untersucht werden sollte.

Bei der Entwicklung einer Softwareanwendung, die Ausnahmen von der Regel entdecken soll, ist es das Ziel, dass das Programm zunächst alle Regeln (Muster) in Bezug auf einen bestimmten Datensatz ermittelt. Das heißt, dass die Software ein "Reverse Engineering" der Regeln durchführen soll, mit denen die Daten erstellt wurden. Dies ist die besondere Stärke von *WizRule*. *WizRule* basiert auf mathematischen Algorithmen, mit denen es möglich ist, alle Regeln, die einen Datensatz bestimmen, innerhalb kürzester Zeit aufzudecken. Das Ergebnis einer *WizRule*-Analyse bildet eine Liste von Datensätzen, die Auffälligkeiten in Bezug auf die ermittelten Regeln aufweisen. Sie deuten auf Verdachtsfälle oder zumindest auf zu untersuchende Fälle hin.

Bei der Nutzung von *WizRule* werden einfach die Daten angegeben, die analysiert werden sollen. Die Software erledigt den Rest. Innerhalb weniger Augenblicke werden die Analyseberichte ausgegeben.

WizRule führt im Laufe einer Datenanalyse folgende Schritte aus:

Im ersten Schritt werden die Daten eingelesen. *WizRule* unterstützt alle gängigen Datenbanksysteme. Im Anschluss haben Sie die Möglichkeit, die Analyseparameter wie „Mindestwahrscheinlichkeit von Wenn-Dann-Regeln“ und „Mindestanzahl von Fällen einer Regel“ zu verfeinern. Sie können genau definieren, nach welchen Arten von Regeln *WizRule* suchen und ob bestimmte Felder ignoriert werden sollen.

WizRule deckt innerhalb kurzer Zeit die Regeln auf, die die Daten bestimmen, und weist auf Fälle hin, die von den ermittelten Regeln abweichen. Jede Abweichung wird zusammen mit den Regeln, von denen sie differiert, ausgegeben.

WELCHE ARTEN VON REGELN UNTERSUCHT *WIZRULE*?

WizRule untersucht den Datenbestand zur Aufdeckung fünf verschiedene Arten von Regeln:

- Formel-Regeln (Formular rules)
- Wenn-Dann-Regeln (If-then rules)
- Wenn-Dann-Formel-Regeln (If-then formular rules)
- Nicht erklärbare Regeln (Outstanding rules)
- Regeln zur Schreibweise (Spelling rules)

Wie bereits erläutert, werden Regeln nicht vom Benutzer vorgegeben. Vielmehr werden alle Regeln automatisch von *WizRule* erkannt.

Ein Beispiel für eine *Formel-Regel* (Formular rule) lautet:

$$\mathbf{A = B * C}$$

Dabei gilt: A = Summe
B = Menge
C = Stückpreis

Genauigkeitsgrad der Regel (Rule's accuracy level): 0,99
Die Regel existiert in 1.890 Datensätzen.

Der „Genauigkeitsgrad“ (Accuracy Level) in Formel-Regeln gibt das Verhältnis zwischen der Anzahl an Fällen, in denen die Formel gilt, und der Gesamtzahl der relevanten Fälle an. Die Werte der Fälle, in denen die Formel gültig ist, können von den exakten Werten geringfügig abweichen (vergleichbar mit dem Toleranzkonzept im Maschinenbau).

WizRule deckt arithmetische Formeln mit bis zu fünf Variablen auf, die in den Daten enthalten sind. Formeln, in denen **A** 0 oder 1 ist, werden ignoriert.

Wenn eine Formel-Regel mit Ausnahme weniger Datensätze für alle Datensätze gilt, sollten die Abweichungen untersucht werden.

Ein Beispiel für eine *Wenn-Dann-Regel* (If-then rule) lautet:

*Wenn **Kundenname** ist Meier*
*und **Artikel** ist Computertyp A*
Dann
Preis = 765

Regelwahrscheinlichkeit (Rule's probability): 0,998
Die Regel existiert in 1.002 Datensätzen.
Signifikanzniveau (Significance level): Fehlerwahrscheinlichkeit < 0,001

Die "*Wahrscheinlichkeit*" (Probability) in Wenn-Dann-Regeln bestimmt das Verhältnis zwischen der Anzahl an Datensätzen, in denen die Bedingung(en) und das Ergebnis enthalten sind, sowie der zugehörigen Anzahl an Datensätzen, in denen die Bedingung(en) mit oder ohne Ergebnis vorliegen.

Das "*Signifikanzniveau*" (Significance level) gibt den Gültigkeitsgrad der Regel an. Es ist gleich 1 minus der "Fehlerwahrscheinlichkeit" (Error probability). Letztere gibt die Wahrscheinlichkeit an, dass die Regel zufällig in den zu analysierenden Daten vorhanden ist.

WizRule deckt alle Wenn-Dann-Regeln mit einer beliebigen Anzahl von Bedingungen auf.

Eine Abweichung von einer als hochgradig gültig eingestuften Regel könnte wiederum auf einen Betrug hindeuten.

Ein Beispiel für eine *Wenn-Dann-Formel-Regel* (If-then formular rule) lautet:

Wenn **Kundenname** ist **Meier**
Dann
Summe = Stückpreis * 0,8 + 50

Genauigkeitsgrad: 0,95
Die Regel existiert in 502 Datensätzen.

Der „*Genauigkeitsgrad*“ (Accuracy Level) gibt den Anteil der Fälle an, die der Formel entsprechen (zuzüglich oder abzüglich einer vorgegebenen Grenze), und zwar bezogen auf die Gesamtanzahl der Fälle, für die die Bedingung gilt.

Eine Abweichung von einer als hochgradig gültig eingestuften Regel könnte wiederum auf einen Betrug hindeuten.

Ein Beispiel für eine *nicht erklärbare Regel* (Outstanding rule) lautet:

Wenn **Kundenname** ist **Meier**
Dann
% Rabatt = 20

Die Regel ist als nicht erklärbar 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 **% Rabatt** korrespondieren.

Mit anderen Worten, die Regel ist nicht erklärbar, da es sich um die einzige Regel handelt, die sich auf einen bestimmten Kunden und einen bestimmten prozentualen Rabatt bezieht (während alle anderen Kunden unterschiedliche Preisnachlässe aufweisen).

Sofern der prozentuale Rabatt aller anderen Kunden zum Beispiel 10% betragen würde, wäre die oben genannte Regel noch erklärbar (da der Rabatt dieses Kunden vom Rabatt aller anderen Kunden abweicht).

Ein Beispiel für eine *Regel zur Schreibweise* (Spelling rule) lautet:

*Der Eintrag **Meier** ist 2.080 Mal vorhanden*
*im Feld **Kundenname**.*

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

Die zugehörigen Regeln werden vorrangig genutzt, um Fälle von abweichend geschriebenen Namen aufzuzeigen. Ein Name wird als abweichend geschrieben angenommen, wenn (a) er einem anderen Namen in diesem Feld ähnelt und (b) die Häufigkeit des ersten Namens (hier Maier) sehr niedrig ist, während die Häufigkeit des zweiten Namens (hier Meier) sehr hoch ist. Wenn beispielsweise der Name Maier nur einmal (im Feld Kundenname) vorhanden ist, wird er als zu untersuchende Abweichung dargestellt.

WIE VERMEIDET *WIZRULE* EINEN FEHLALARM?

Nach der Feststellung der Regeln, die dem zu untersuchenden Datenbestand zugrunde liegen, prüft *WizRule* auf Abweichungen von diesen Regeln. Dabei muss jedoch nicht jede Regelabweichung geprüft werden. Nehmen wir an, *WizRule* deckt folgende *Wenn-Dann-Regel* auf:

Wenn Kundename ist Meier

Dann

Verkäufer ist Schulze

Regelwahrscheinlichkeit (Rule's probability): 0,98

Die Regel existiert in 1.003 Datensätzen.

Signifikanzniveau (Significance level): Fehlerwahrscheinlichkeit < 0,001

Da die Regelwahrscheinlichkeit 0,98 beträgt und die Regel in 1.003 Datensätzen vorhanden ist, gibt es ungefähr zwanzig Datensätze, in denen der Verkäufer von dieser Regel abweicht. Die Überprüfung aller zwanzig Datensätze ist langwierig. Eine Vielzahl der Abweichungen können einen Fehlalarm darstellen.

Um derartige Fehlalarme zu vermeiden, prüft *WizRule*, ob die Abweichung durch eine andere, in den Daten enthaltene Regel erklärbar ist. Sofern dies der Fall ist, wird keine Auffälligkeit ausgewiesen. Wenn es sich bei dem Artikel beispielsweise um einen Computer handelt, ist der Verkäufer *Müller*. Auf diese Weise kann die Regel einige der oben aufgeführten Abweichungen begründen. Sofern Abweichungen erklärbar sind, werden sie nicht als zu untersuchende Fälle betrachtet.

WizRule prüft weiterhin die Häufigkeit des *Dann*-Werts im Abweichungsfall. Seine Häufigkeit sollte in Bezug auf die Regelbedingungen niedriger sein als seine Gesamthäufigkeit im Datenbestand. Wenn nicht, wird der Fall wiederum nicht als zu untersuchender Fall betrachtet. Beispielsweise wenn der Verkäufersname in zwei der abgelehnten Datensätze *Schmidt* lautet und dieser in nur zwei Fällen im gesamten Datenbestand vorhanden ist, werden diese Fälle nicht als zu untersuchende Fälle betrachtet. Ist *Schmidt* jedoch für gewöhnlich der Verkäufer anderer Kunden, so handelt es sich in der Tat um Verdachtsfälle.

Sofern das Feld *Dann* numerisch ist, kann *WizRule* Fehlalarme zu reduzieren, indem nur die Abweichungen ausgegeben werden, bei denen der *Dann*-Wert vom erwarteten Wert um mindestens eine Standardabweichung differiert. Kleinere Abweichungen werden ignoriert.

Bei der Betrachtung verdächtiger Fälle können die Fälle über das Feld *Dann* sortiert und die Werte innerhalb des Feldes wiederum sortiert werden. Im Falle einer Vielzahl von Abweichungen ist die Sortierung der Abweichungen eine effiziente Methode für die Fokussierung auf die auffälligsten Fälle.

WizRule vermeidet fast alle Fehlalarme aufgrund der Berücksichtigung der differenzierten Methodik. Es werden nur die zu untersuchenden Fälle gelistet.

IST *WIZRULE* NUR FÜR MATHEMATIKER GEEIGNET?

Obwohl *WizRule* auf komplexen mathematischen Algorithmen basiert, wurde die Software für Benutzer entwickelt, die über geringe oder keine mathematischen Kenntnisse verfügen. *WizRule* führt Berechnungen im Hintergrund durch und stellt die Ergebnisse der Analyse in übersichtlichen, leicht verständlichen Formaten dar.

Selektieren Sie einfach die zu analysierende Datei - *WizRule* erledigt den Rest der Analyse für Sie.

Wenn Sie jedoch an der Mathematik und Berechnungsmethoden hinter *WizRule* interessiert sind, verweisen wir auf folgendes Dokument auf unserer Website:

<https://www.wizsoft.com/index.php/products/WizRule/WizRule-technologies/>

###

WizSoft Inc. • 1 (516) 393 5841 • www.wizsoft.com • info@wizsoft.com

WizRule und WizSoft sind eingetragene Marken von WizSoft, Inc.
Bei weiteren aufgeführten Namen kann es sich um Marken anderer Hersteller handeln.